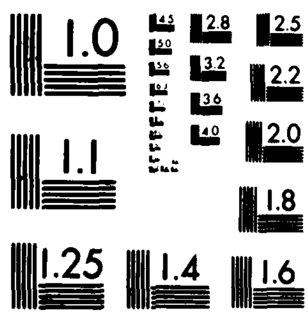
AD-A110 816 NAVAL POSTGRADUATE SCHOOL MONTEREY CA F/G 4/2
A STATISTICAL ANALYSIS OF MONTHLY RAINFALL FOR MONTEREY PENINSU--ETC(U)
MAR 81 D F DAVIS
UNCLASSIFIED NL

1 of 3
AD
A110816


1.0
2.8
2.5
3.2
2.2
3.6
4.0
2.0
1.1
1.8
1.25
1.4
1.6

LEVEL II

AD A110816

NAVAL POSTGRADUATE SCHOOL
Monterey, California

DTIC ELECTE FEB 11 1982 H

THESIS

A STATISTICAL ANALYSIS OF MONTHLY RAINFALL FOR MONTEREY PENINSULA AND THE CARMEL VALLEY IN CENTRAL CALIFORNIA

by

David Frederick Davis

March 1981

Thesis Advisor: P.A. Jacobs

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER	2. GOVT ACCESSION NO. AD-A110
3. RECIPIENT'S CATALOG NUMBER 816	
4. TITLE (and Subtitle) A Statistical Analysis of Monthly Rainfall for Monterey Peninsula and the Carmel Valley in Central California	5. TYPE OF REPORT & PERIOD COVERED Master's Thesis; March 1981
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) David Frederick Davis	8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Postgraduate School Monterey, California 93940	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Postgraduate School Monterey, California	12. REPORT DATE March 1981
	13. NUMBER OF PAGES 207
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report) Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Monthly Rainfall Models
California Rainfall
Seasonal Rainfall Forecasting
ARMA Models
Logistic Analysis
Logistic Time Series

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This thesis presents a statistical analysis of the monthly rainfall for the Monterey Peninsula and the Carmel Valley in Central California. The analysis begins with the simple first-order autoregressive Markov model, which is found to be weak. Next, 2x2 contingency tables are used to identify predictors, one of which is found to be January rainfall. Finally, logistic analysis is used to quantify the predictive ability

DD FORM 1473, 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

of January.

This paper attempts to analyze rainfall time series in the statistical sense. No attempt is made to provide a physical explanation of the findings from the point of view of a meteorologist.

Accession For	
NTIS GRA&I	✓
DTIC TAB	
Unannounced	
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A	

Approved for public release; distribution unlimited

A Statistical Analysis of Monthly Rainfall for Monterey Peninsula and the Carmel Valley in Central California

by

David Frederick Davis
Captain, United States Army
B.S., Colorado School of Mines, 1972

Submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
March 1981

Author: ______________________

Approved by: ______________________
Thesis Advisor

____
Second Reader

____
Chairman, Department of Operations Research

W M Woods
Dean of Information and Policy Sciences

ABSTRACT

This thesis presents a statistical analysis of the monthly rainfall for the Monterey Peninsula and the Carmel Valley in Central California. The analysis begins with the simple first-order autoregressive Markov model, which is found to be weak. Next, 2x2 contingency tables are used to identify predictors, one of which is found to be January rainfall. Finally, logistic analysis is used to quantify the predictive ability of January.

This paper attempts to analyze rainfall time series in the statistical sense. No attempt is made to provide a physical explanation of the findings from the point of view of a meteorologist.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENT

I would like to thank my advisors, Professors D. P. Gaver and P. A. Jacobs for their guidance. My second reader, Professor R. J. Renard has made sure that no laws of meteorology have been destroyed, for which I am grateful.

Other who assisted me should also be mentioned, especially Professor A. L. Schoenstadt who allowed me access to the HP 9845B computer which drew all the figures. The personnel at the Monterey Peninsual water Management District, Dr. J. Williams, Mr. B. Buel, and Mr. K. Walsh, were also very helpful.

I. INTRODUCTION

A. THE PROBLEM

The Monterey Peninsula Water Management District, in Central California coastal area has as one of its responsibilities the duty to recommend and/or impose water rationing on its constituents. To do this in a rational way requires the District to have some formula for predicting future water availability. Although the techniques of modern meteorology are becoming more sophisticated and exact there is still the inability to make good long-range predictions. This thesis analyzes three series of Monterey County monthly rainfall data by purely statistical methodology in order to identify possible predictive formulas.

B. NOTATION

Rainfall will be denoted as $R_{t,m}$ which will represent inches of rain recorded for the t^{th} year and the m^{th} month. The year to be used is the California Water Year which begins in October and ends the following September. Thus $R_{1,1}$ is the monthly rainfall for October of year '1' and $R_{6,8}$ is the monthly rainfall of May of year '6'.

An overstruck bar as in $\bar{R}_{..}$ will indicate the arithmetic average of a variable; in this case it is the arithmetic average of all years and months of rainfall. $\bar{R}_{.m}$ is the average of rainfall over the years for month m; $\bar{R}_{t.}$

represents the yearly average for year t.

C. METHODS OF ANALYSIS

Three methods were used to analyze the data. The first method was to model the series using autoregressive moving averages as described in Box and Jenkins [Ref. 1]. The second was to use 2x2 contingency tables to identify possible predictors. The third was logistic regression to quantify the findings of the 2x2 contingency table analysis. These three methods will be described in further sections of this paper.

1. ARMA(p,q) Models

A widely used approach to time series modeling proposed by Box and Jenkins is the ARMA(p,q) model. This model is actually a joining of two types of model, the autoregressive and the moving average.

In the noation of Box and Jenkins:
let $\{Z_t, t=1,2,\ldots,n\}$ be a time series, then an ARMA(p,q) process may be written as:

$$\tilde{Z}_t = \phi_1 \tilde{Z}_{t-1} + \cdots + \phi_p \tilde{Z}_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q} \qquad \text{I.1}$$

the $\{a_t, t=1,2,\ldots,n\}$ are assumed to be random shocks distributed as independent and identically distributed (iid) random variables with mean zero and variance σ_a^2 and $\tilde{Z}_t = Z_t - \overline{Z}..$ The further assumption of normality is also usually made.

For purposes of this paper, a mapping of $R_{t,m}$ into Z_r, $r=12(t-1)+m$ was made, and an ARMA analysis was

then conducted on this index transformed series. This analysis is described in section III.

2. <u>2x2 Table Analysis</u>

In the validation of section IV it is found that the ARMA model is not very successful in describing the data. In section V the data is analyzed by means of 2x2 contingency tables. These tables are good tools for exploratory data analysis in that they provide a visual display of the data. Statistical procedures based on the null hypothesis of independence can be used to quantify the departure from independence. The theory of 2x2 tables, and contingency tables in general may be found in Fleiss [Ref. 3], Dixon and Massey [Ref. 5], Brownlee [Ref. 6], and Mood, Graybill, and Boes [Ref. 7].

For this paper, the contingency table approach is used to identify a month or group of months of a year whose rainfall can serve as a predictor for the rainfall during the remaining months of the year. One predictor that was suggested is the rainfall in the month of January.

3. <u>Logistic Analysis</u>

Once a predictor is tentatively identified it becomes necessary to quantify the degree, direction and accuracy of the predictor.

A logistic analysis is conducted by dividing the data for a year into two sets, the predictor or control set, and predictand or complement set. For this analysis, the predictor

is the logged anomaly of January rainfall for the year; that is, if X_t denotes the predictor or control for year t, then

$$X_t = \ell n(R_{t,4}) - \frac{1}{N} \sum_{t=1}^{N} \ell n(R_{t,4}) \qquad \text{I.2}$$

(The logarithm is used to better symmetrize the model.) The complement is the raw anomaly of the total rainfall for the immediately subsequent eleven months; that is, if Y_t' denotes the complement for year t, then;

$$Y_t' = \left(\sum_{m=5}^{12} R_{t,m} + \sum_{m=1}^{3} R_{t+1,m}\right) - \frac{1}{N-1} \sum_{t=1}^{N-1} \left(\sum_{m=5}^{12} R_{t,m} + \sum_{m=1}^{3} R_{t+1,m}\right) \qquad \text{I.3}$$

Finally, the data are transformed into a binary representation, relative to zero as;

$$Y_t = \begin{cases} 0 & \text{if } Y'_t < 0 \\ 1 & \text{if } Y'_t > 0 \end{cases} \qquad \text{I.4}$$

In section VI the model fit is

$$P(Y=1 | X=x) = \frac{e^{\alpha+\beta x}}{1 + e^{\alpha+\beta x}} \qquad \text{I.5}$$

Where x is as before and $P(Y=1|X=x)$ is interpreted as: "the conditional probability that the subsequent eleven month total rainfall will be above its mean, given that the logged anomaly of January rainfall was 'x'".

II. THE DATA

A. GENERAL

Three data sets were used for this analyis. The location at which these data sets were gathered is shown in Figure 1. As the figure indicates, two of the data sets are on the Monterey Peninsula proper, while the third set, SC, represents the Carmel River Watershed at the San Clemente Dam.

Although data exists in all cases to the present, all three sets were truncated at September of 1974. The remaining data, up through September of 1980 was reserved for validation of the models and methodology.

The data coordinates are:

Data set RN:	36° 35' 42" North Latitude 121° 54' 43" West Longitude
Data set FL:	36° 35' 30" North Latitude 121° 56' 30" West Longitude
Data set SC:	36° 26' 12" North Latitude 121° 42' 30" West Longitude

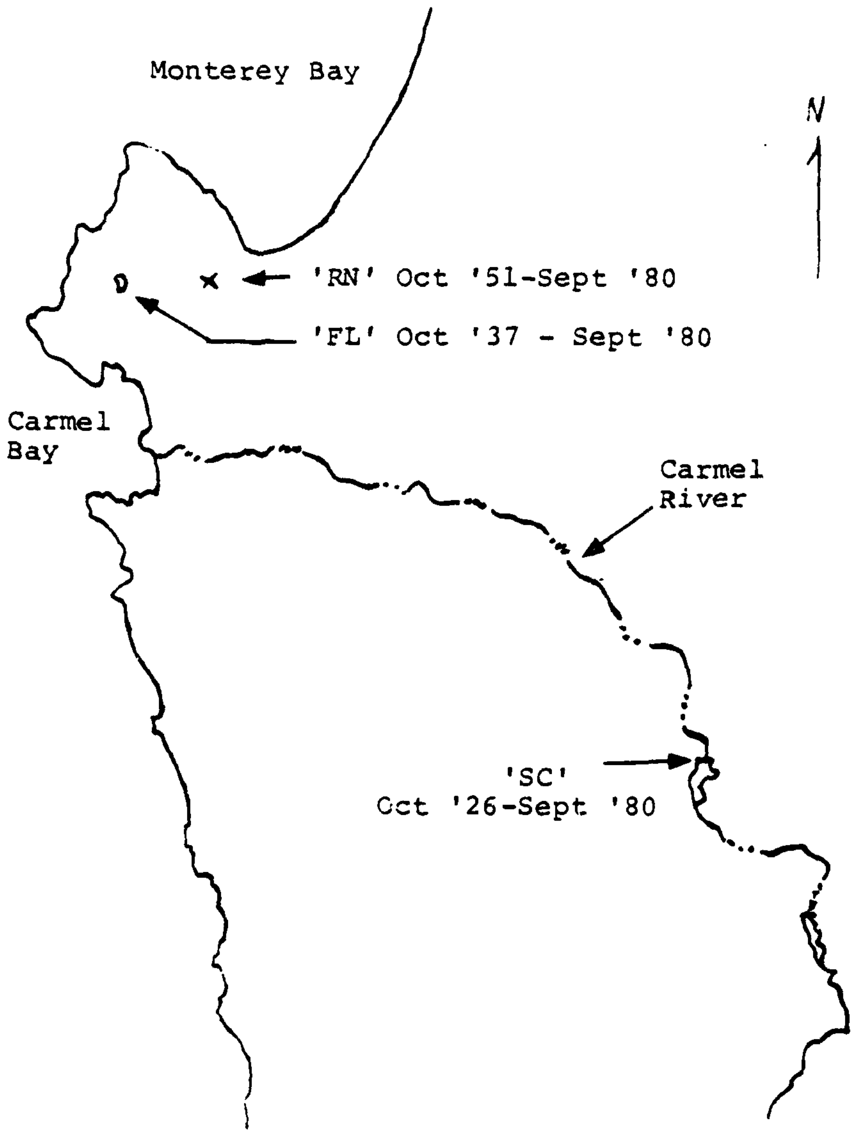


Figure 1. Location of rainfall data sets and the years available.

B. DATA SET RN

Data set RN consists of monthly rainfall amounts gathered by Professor R.J. Renard, Cooperative Observer for the National Weather Service Climatological Station, Monterey, California. The data set begins in June 1951 and currently terminates in September 1980. As was stated above, the analysis was conducted only on that data between and including October 1951 and September 1974.

1. Raw Data

Appendix A contains a listing of data set RN. Figure 2 shows the raw data set. Month 1 is October 1951, month 148 is January 1964, and up to month 288 which is September 1974. As can be seen the data are strongly seasonal. This is enough to indicate that the series, as stated, is highly non-stationary.

The data presented so far deals with only monthly data. Next to be considered is the series of yearly total rainfalls. The results are shown in Figure 3 (Yearly total rainfall), 4 (Correlogram of yearly rainfall), and Table 1 (Estimated Autocorrelations). In this case, the correlogram indicates stationarity and independence of the yearly series. A plot of the lag one relationships, Figure 5, reinforces this indication of independence.

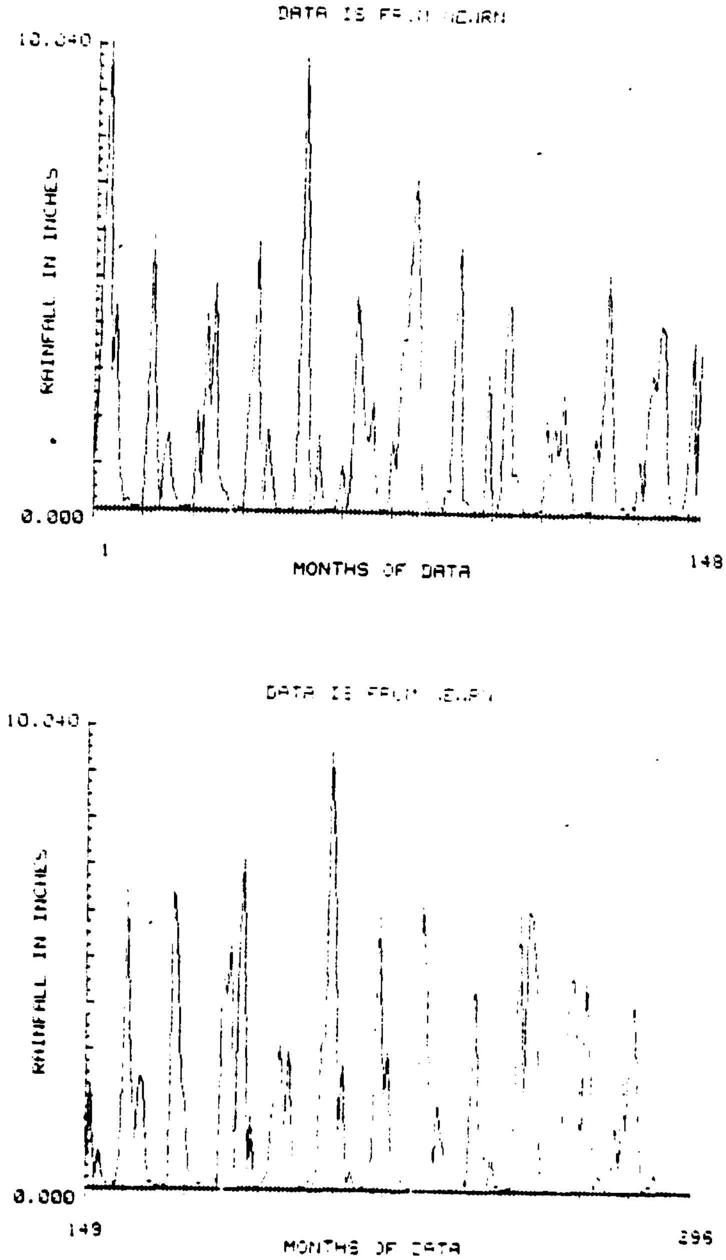


Figure 2. Monthly rainfall in inches for data set RN.

The correlograms and Partial Correlograms to follow indicate the 95% approximate significance levels using dashed lines. For development of these significance levels see Box and Jenkins [Ref. 1].

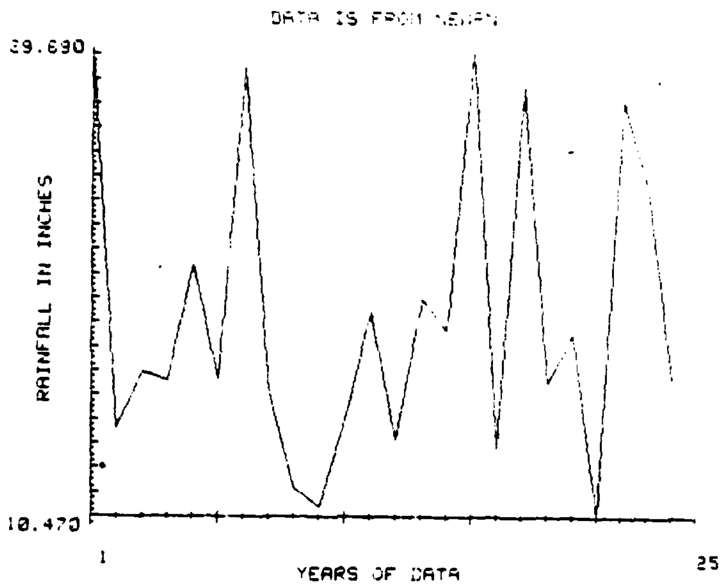


Figure 3. Yearly total rainfall for data set RN (1951 - 1974).

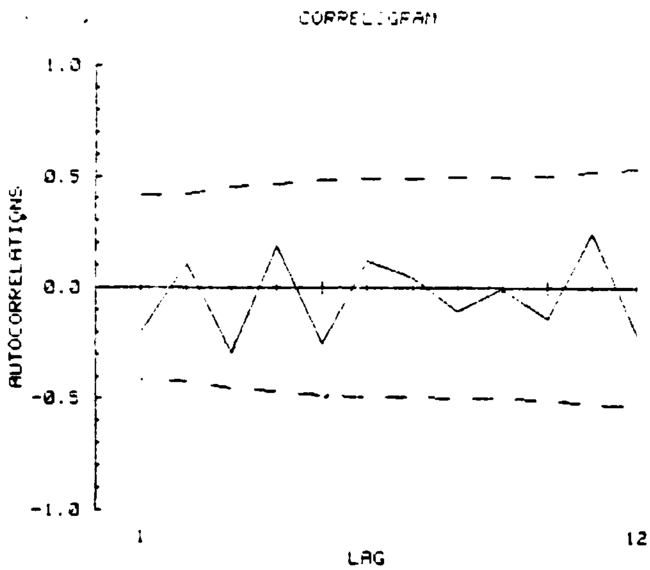


Figure 4. Correlogram of Yearly total rainfall for data set RN.

TABLE 1

ESTIMATED AUTOCORRELATIONS OF YEARLY TOTAL RAINFALL FOR DATA SET RN

AUTOCORRELATIONS

LAG	VALUE	LAG	VALUE
1	-.200	7	.044
2	.109	8	-.109
3	-.295	9	-.009
4	.186	10	-.139
5	-.249	11	.248
		12	-.211

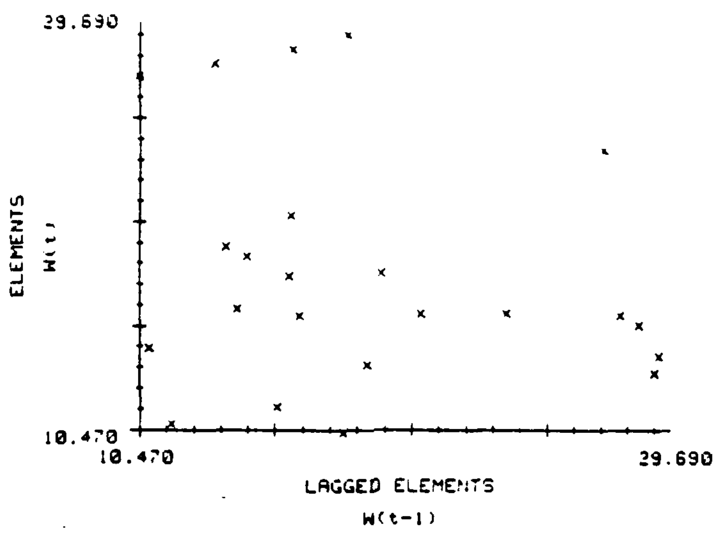


Figure 5. Lag one plot of yearly rainfall data for data set RN.

2. Swept Data

Pierce [Ref. 9] and Hipel [Ref. 11] suggest various ways to remove the seasonality of data sets like RN, FL, and SC. The basic, and most straight forward of these methods is to remove the various monthly means. This is accomplished by the following replacement:

let

$$\tilde{R}_{t,m} = R_{t,m} - \bar{R}_{\cdot m} \qquad \text{II.5}$$

where $\bar{R}_{\cdot m}$ represents the mean of the month m.

One statistic that is a byproduct of the calculations of $\bar{R}_{\cdot m}$ is $S^2_{\cdot m}$ defined as the estimated variance of the

monthly data points:

$$S^2_{\cdot m} = \frac{1}{N-1} \left(\sum_{t=1}^{N} R^2_{t,m} - N\bar{R}_{\cdot m}^{2} \right); \qquad \text{II.6}$$

These statistics for data set RN are shown in Table 2, and illustrated in Figure 6. In the same way as the raw data mapped into a series, a series is created from $\tilde{R}_{t,m}$ as;

$$\tilde{z}_r = \tilde{R}_{t,m} \quad ,r=12(t-1)+m \qquad \text{II.7}$$

TABLE 2

MONTHLY MEANS AND VARIANCE FOR DATA SET RN

MONTH	MEAN	VARIANCE
1	.677	.5292
2	2.478	3.6354
3	3.355	5.9147
4	4.124	5.4146
5	2.592	4.5122
6	2.639	3.7414
7	1.708	2.7699
8	.434	.2494
9	.219	.1004
10	.058	.0056
11	.104	.0129
12	.275	.3953

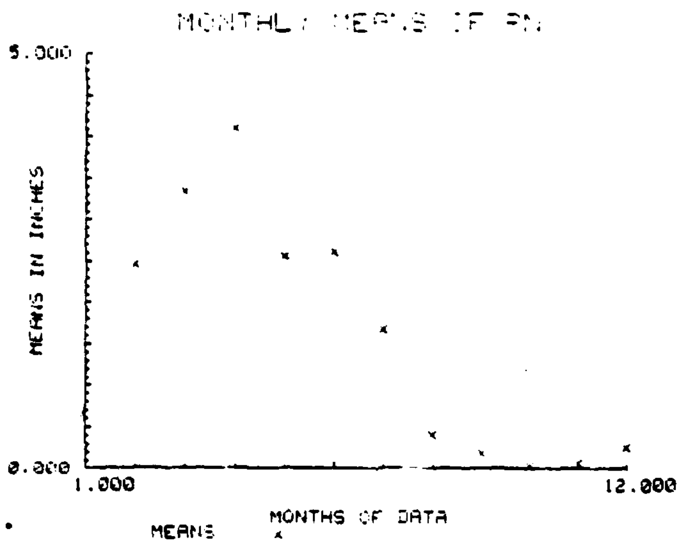


Figure 6. Monthly means for data set RN

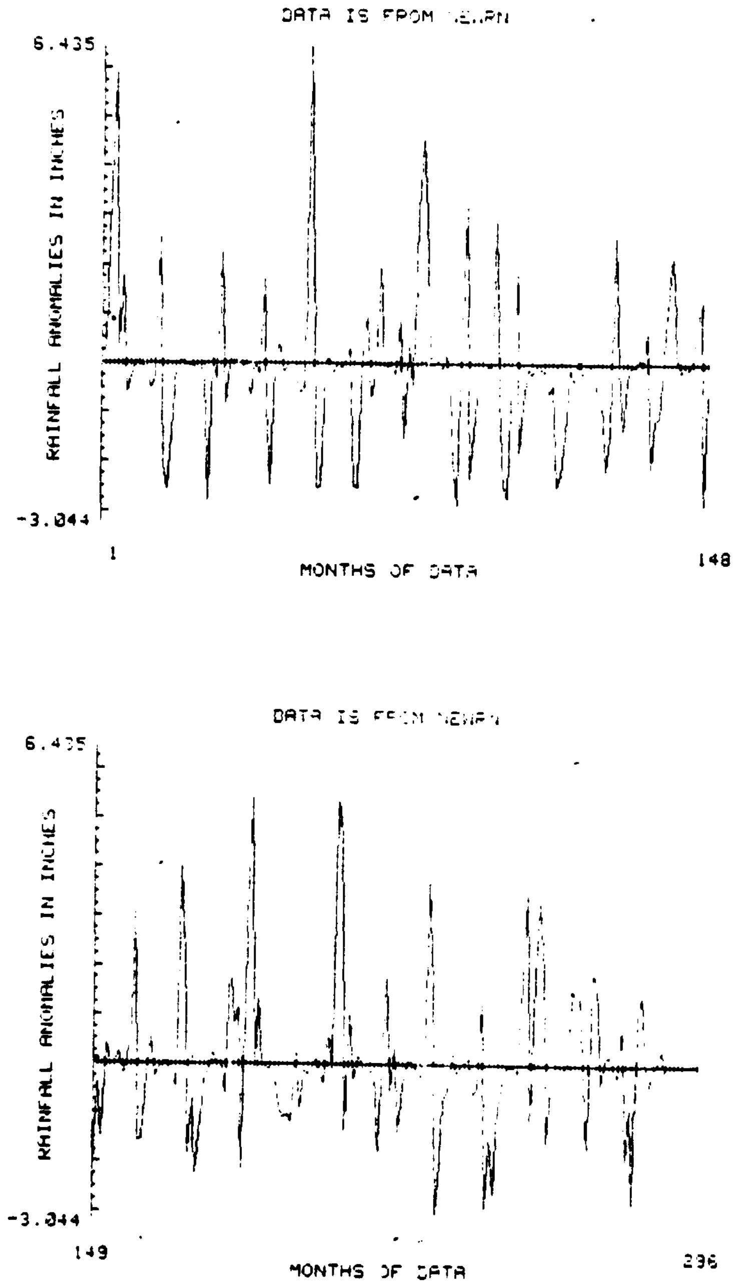


Figure 7. Monthly rainfall anomalies in inches for data set RN

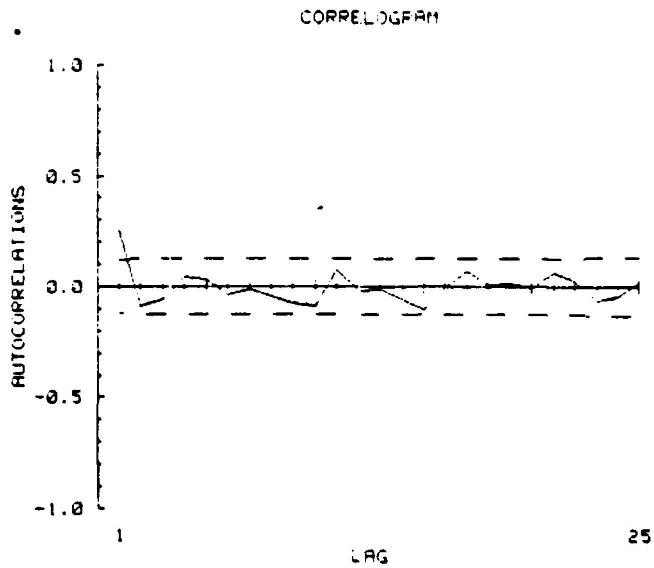


Figure 8. Correlogram of the monthly rainfall anomalies for data set RN

TABLE 3

ESTIMATED AUTOCORRELATIONS OF MONTHLY RAINFALL ANOMALIES FOR DATA SET RN

LAG	VALUE	LAG	VALUE
1	.249	14	-.057
2	-.090	15	-.109
3	-.059	16	-.006
4	.041	17	.067
5	.032	18	.004
6	-.035	19	.013
7	-.011	20	-.007
8	-.043	21	.063
9	-.073	22	.023
10	-.091	23	-.066
11	.076	24	-.044
12	-.020	25	.021
13	-.012		

3. Logged and Swept Data

The data should now be stationary in the means. However, as seen in Table 2, the variances of monthly rainfall amounts are not homogeneous. Kilmartin [Ref. 10] discusses various transformations of the data to remove this heteroskedacity. A plot of the variance versus mean, Figure 9 below, indicates that the logarithmic transform of the data might be useful.

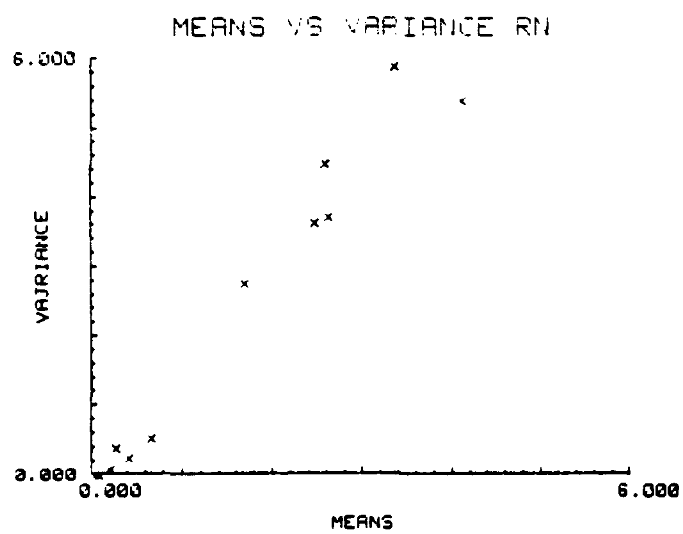


Figure 9. Plot of monthly variance against monthly means for data set RN

Since the data contain zeros, the following modified logarithmic transformation is done

$$R'_{t,m} = \ell n(R_{t,m} + 1) - \frac{1}{N} \sum_{t=1}^{N} \ell n(R_{t,m} + 1) \qquad \text{II.8}$$

where the effect of the addition of the one is mostly to preserve the mapping of zeros into zeros. A more in depth discussion of this transformation is found in Kilmartin. The mapping is performed again as before and $\tilde{R}'_{.m}$ and $S'_{.m}$ are calculated in a manner similar to II.6 and shown in Table 4 and Figures 10 and 11.

TABLE 4

MONTHLY MEANS AND VARIANCE FOR
LOGGED DATA SET RN

MONTH	MEAN	VARIANCE
1	.438	.1549
2	1.092	.3454
3	1.312	.3538
4	1.539	.2003
5	1.104	.3795
6	1.133	.3706
7	.854	.2728
8	.319	.0767
9	.176	.0401
10	.054	.0045
11	.094	.0094
12	.185	.0849

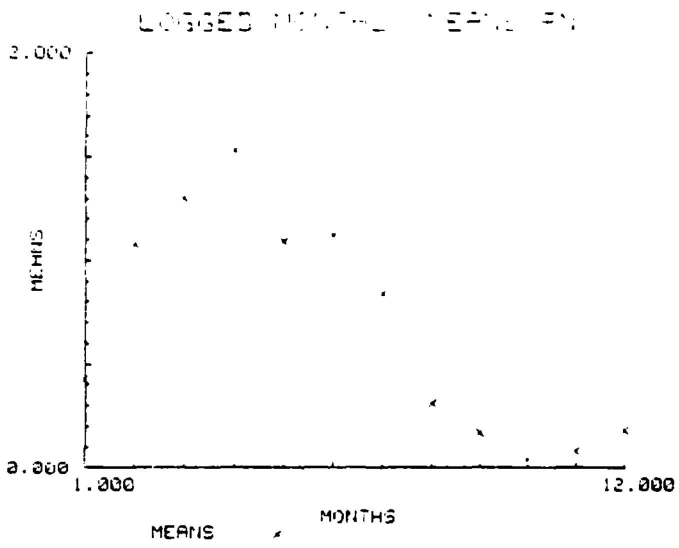


Figure 10. Monthly means of logged data set RN

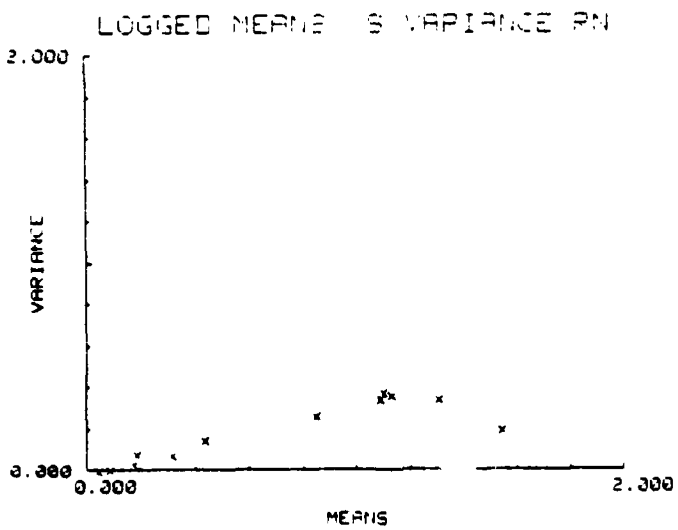


Figure 11. Plot of monthly variance against monthly means for logged data set RN

These transformations, the logarithm followed by the removal of the monthly means of the logged data, result in the series listed in Appendix A and described in Figures 12 and 13 with Table 5.

These displays indicate that a suitably stationary series has been obtained. Other methods, such as differencing, scaling, and Box-Cox transformations, see Hipel [Ref. 11], were tried but with less success.

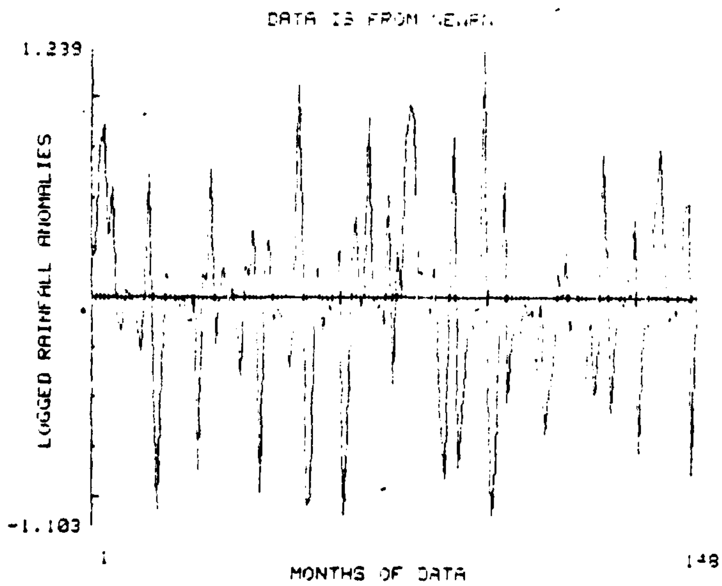


Figure 12a. Logged anomalies of monthly rainfall for data set RN. Months 1 -148

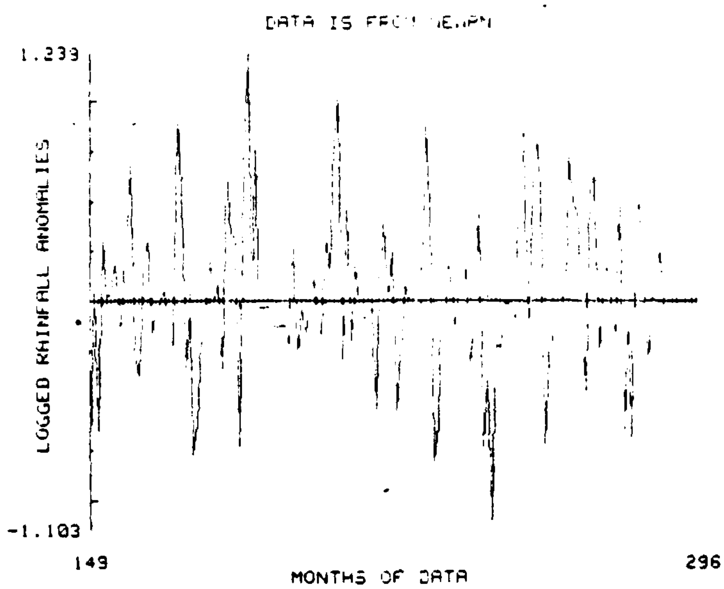


Figure 12b. Logged anomalies of monthly rainfall for data set RN

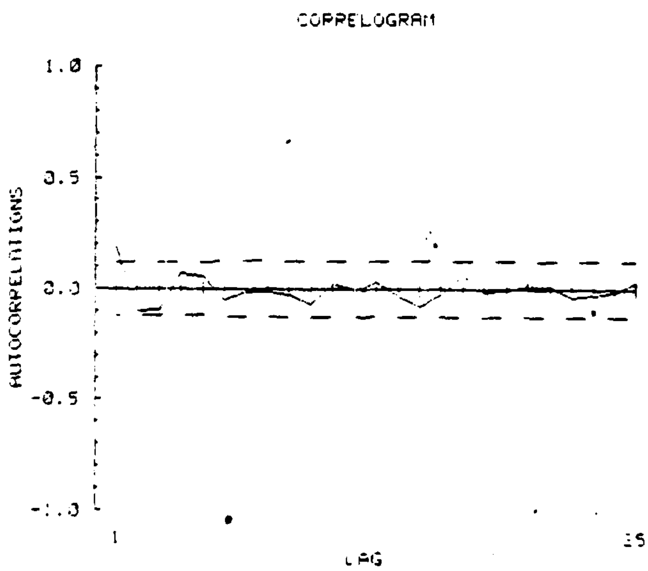


Figure 13. Correlogram of logged anomalies of monthly rainfall from data set RN

TABLE 5

ESTIMATED AUTOCORRELATIONS OF LOGGED ANOMALIES OF MONTHLY RAINFALL FROM DATA SET RN

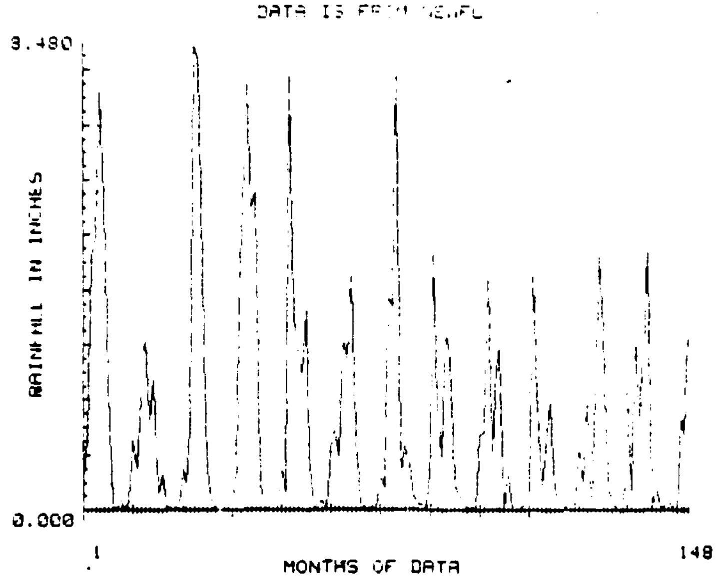
AUTOCORRELATIONS

LAG	VALUE	LAG	VALUE
1	.191	14	-.033
2	-.102	15	-.084
3	-.095	16	-.024
4	.071	17	.062
5	.056	18	-.015
6	-.053	19	-.008
7	-.009	20	.013
8	-.014	21	.013
9	-.022	22	-.038
10	-.069	23	-.024
11	.024	24	-.012
12	-.004	25	.032
13	.032		

C. DATA SET FL

The label for these data derives from its location, Forest Lake, on the Monterey Peninsula, in Pebble Beach, California. Data set FL consists of monthly rainfall figures gathered by the California-American Water Company since 1896. Although this data set started quite early, the data prior to 1937 has frequent missing observations. Therefore, this data set is taken as October 1937 through September 1974, with October 1974 through September 1980 reserved for validation.

Analysis of this data set is identical to that of data set RN, therefore only the pertinent figures and tables are shown.



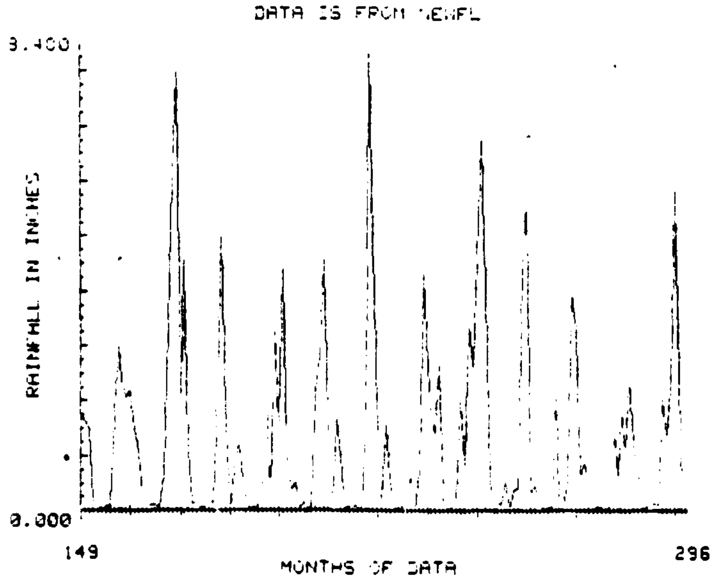


Figure 14a. Months,1 - 296 of rainfall in inches for data set FL

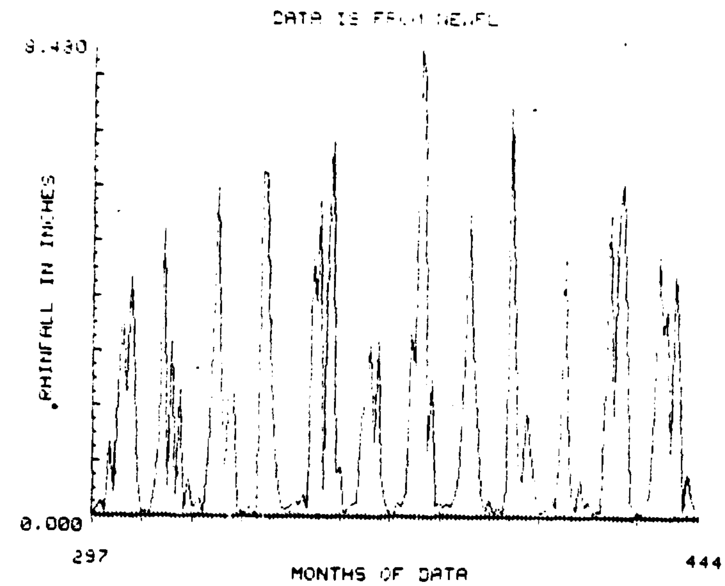


Figure 14b. Months 297-444 of rainfall in inches for data set FL

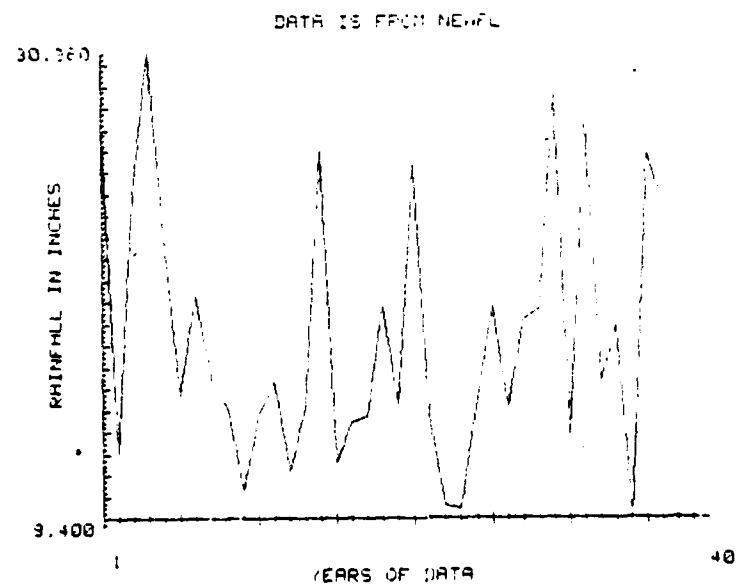


Figure 15. Yearly total rainfall for data set FL (1937 - 1974).

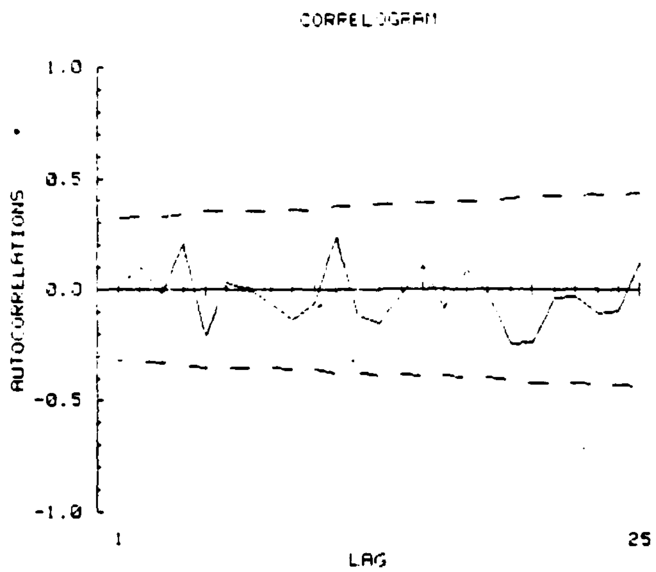


Figure 16. Correlogram of yearly total rainfall for data set FL

TABLE 6

ESTIMATED AUTOCORRELATIONS OF YEARLY TOTAL RAINFALL FOR DATA SET FL

AUTOCORRELATIONS

LAG	VALUE	LAG	VALUE
1	.010	14	-.028
2	.099	15	.105
3	-.022	16	-.087
4	.207	17	.085
5	-.214	18	-.018
6	.028	19	-.245
7	-.003	20	-.237
8	-.061	21	-.035
9	-.138	22	-.031
10	-.060	23	-.106
11	.236	24	-.096
12	-.119	25	.120
13	-.151		

2. Swept Data

TABLE 7

MONTHLY MEANS AND VARIANCE FOR DATA SET FL

MONTH	MEAN	VARIANCE
1	.744	.4895
2	2.235	3.7444
3	3.049	4.2480
4	3.537	4.4240
5	2.999	5.9492
6	2.724	3.1743
7	1.559	2.4505
8	.449	.2033
9	.153	.0417
10	.077	.0081
11	.115	.0081
12	.189	.1350

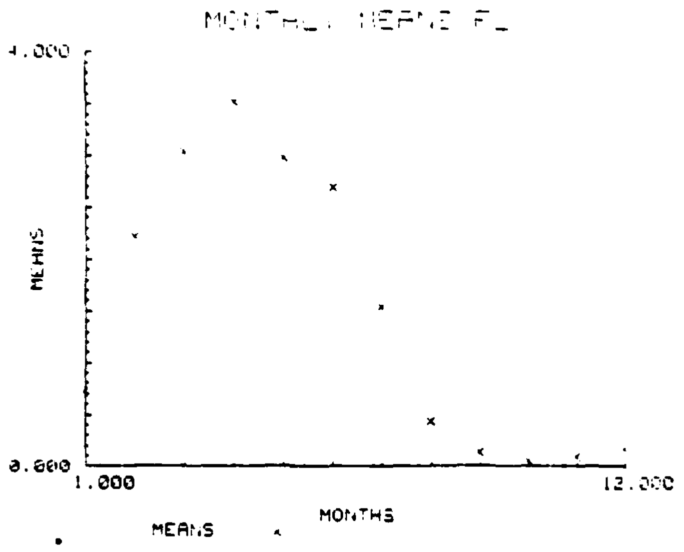
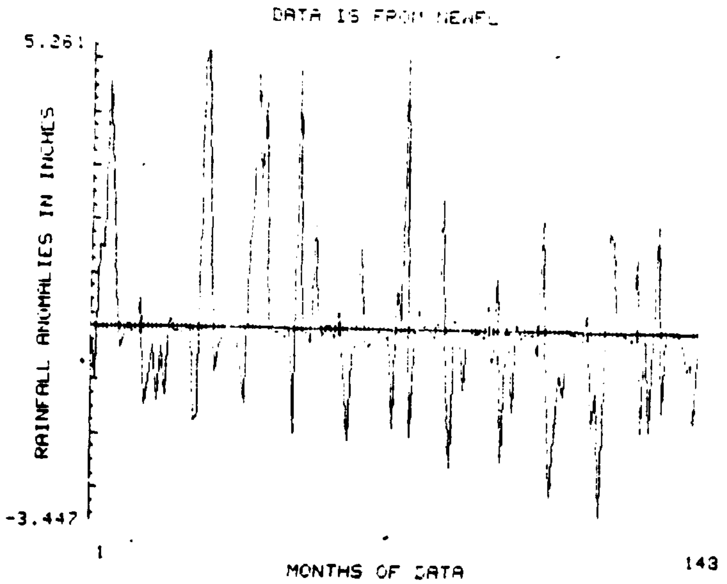


Figure 17. Monthly means for data set FL



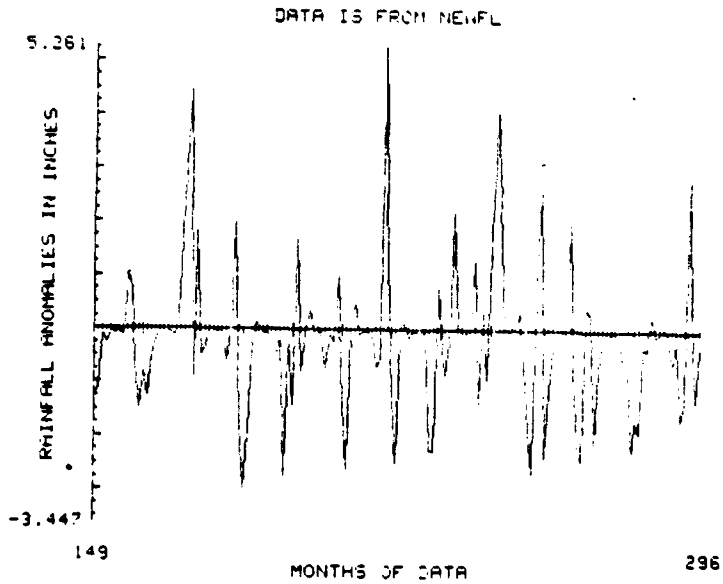


Figure 18a. Months 1 - 296 of rainfall anomalies in inches for data set FL

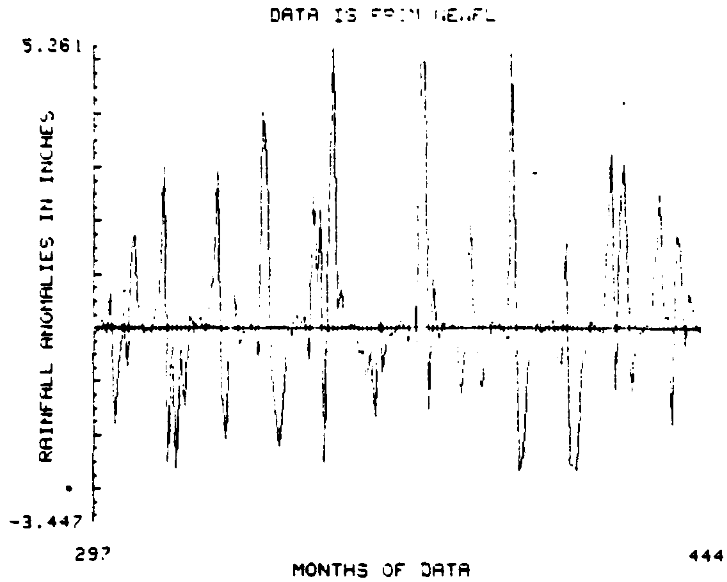


Figure 18b. Months 297-444 of rainfall anomalies in inches for data set FL

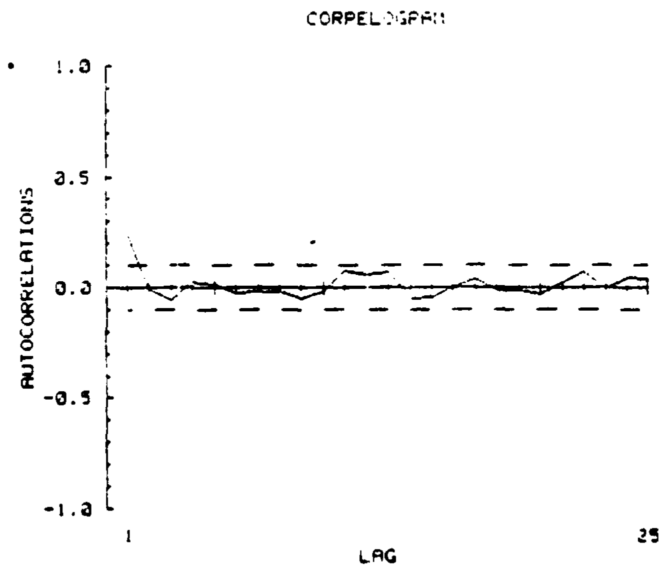


Figure 19. Correlogram of monthly rainfall anomalies for data set FL

TABLE 8

ESTIMATED AUTOCORRELATIONS OF MONTHLY RAINFALL ANOMALIES FOR DATA SET FL

AUTOCORRELATIONS

LAG	VALUE	LAG	VALUE
1	.244	14	-.053
2	-.007	15	-.043
3	-.056	16	-.002
4	.027	17	.039
5	.016	18	-.014
6	-.026	19	-.011
7	-.022	20	-.031
8	-.021	21	.023
9	-.051	22	.067
10	-.020	23	-.007
11	.077	24	.041
12	.059	25	.039
13	.068		

3. <u>Logged and Swept Data</u>

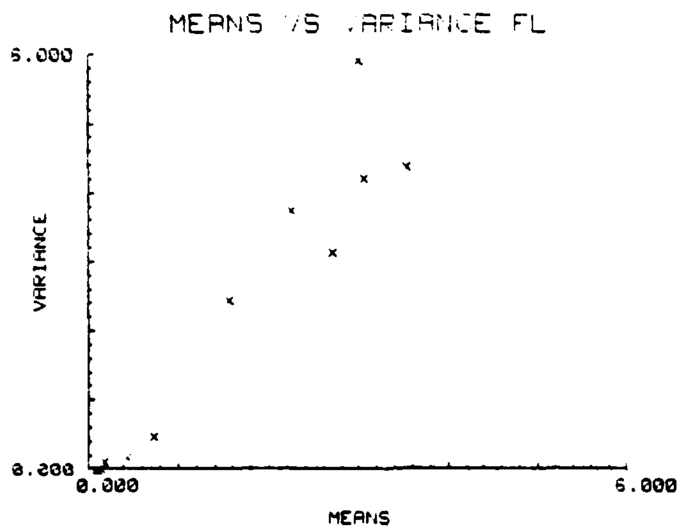


Figure 20. Plot of monthly variance against monthly means for data set FL

TABLE 9

MONTHLY MEANS AND VARIANCE FOR LOGGED DATA SET FL

MONTH	MEAN	VARIANCE
1	.484	.1440
2	1.007	.3440
3	1.269	.2781
4	1.398	.2579
5	1.218	.3421
6	1.184	.3006
7	.796	.2717
8	.338	.0589
9	.130	.0234
10	.071	.0061
11	.106	.0064
12	.146	.0044

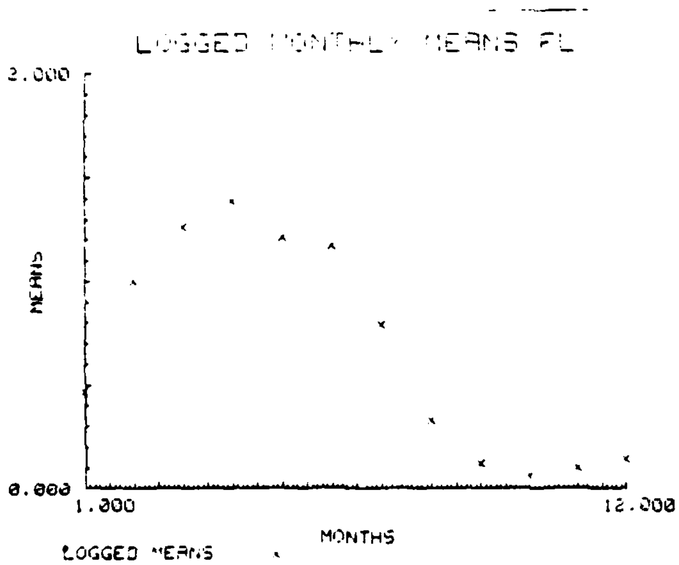


Figure 21. Monthly means of logged data set FL

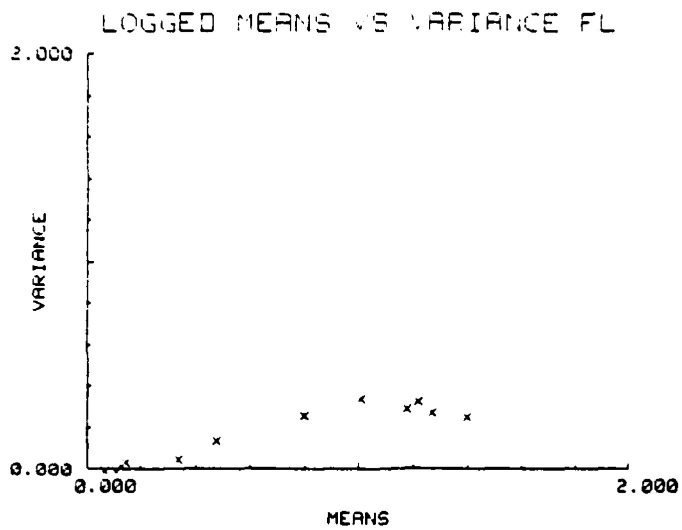


Figure 22. Plot of monthly variance against monthly means for logged data set FL

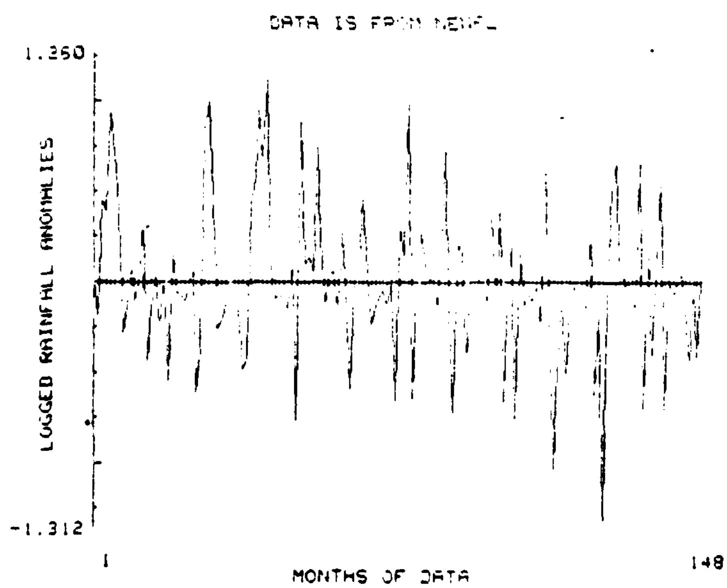
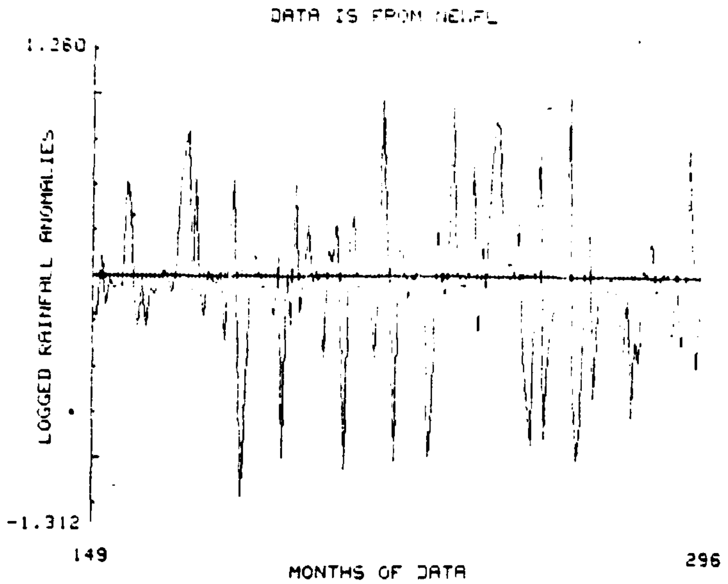


Figure 23a. Months 1 - 148 of logged rainfall anomalies for data set FL



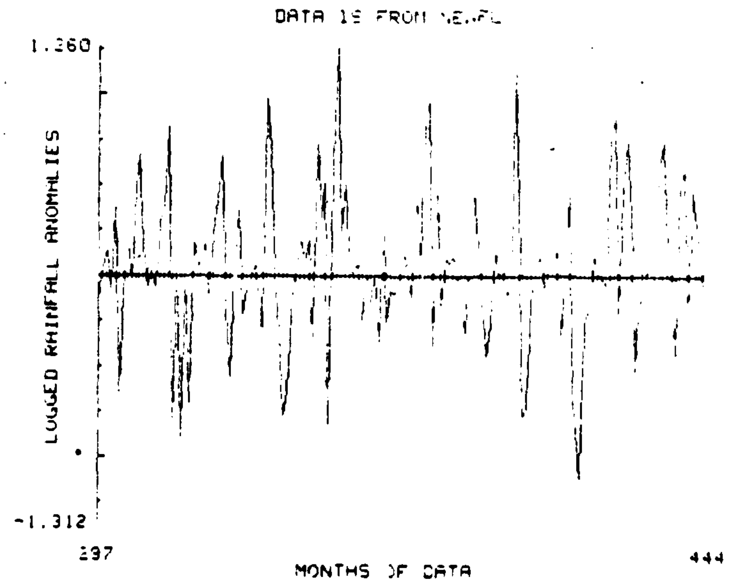


Figure 23b. Months 149 - 444 of logged rainfall anomalies from data set FL

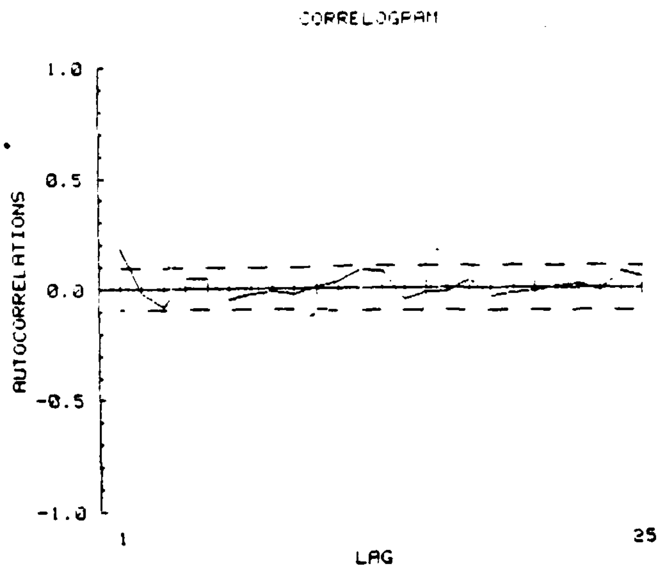


Figure 24. Correlogram of logged anomalies of monthly rainfall from data set FL.

TABLE 10

ESTIMATED AUTOCORRELATIONS OF LOGGED ANOMALIES FROM MONTHLY RAINFALL OF DATA SET FL

AUTOCORRELATIONS

LAG	VALUE	LAG	VALUE
1	.185	14	-.052
2	-.020	15	-.021
3	-.081	16	-.020
4	.046	17	.040
5	.043	18	-.040
6	-.050	19	-.025
7	-.024	20	-.014
8	-.010	21	.007
9	-.027	22	.019
10	.004	23	-.015
11	.031	24	.076
12	.084	25	.047
13	.074		

D. DATA SET SC

The label for this data derives for its location, San Clemente Dam, on the Carmel River in Central Califronia, approximately 26 kilometers southeast of data sets RN and FL on the Monterey Peninsula. Data set SC consists of monthly rainfall figures gathered by the California-American Water Company since 1926.

Analysis of this data set is again very close to that of the previous data sets and only the displays will be given.

1. Raw Data

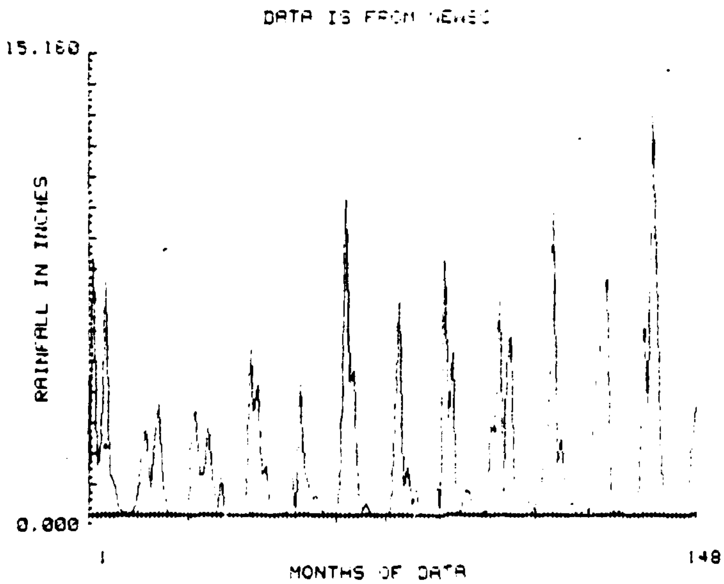
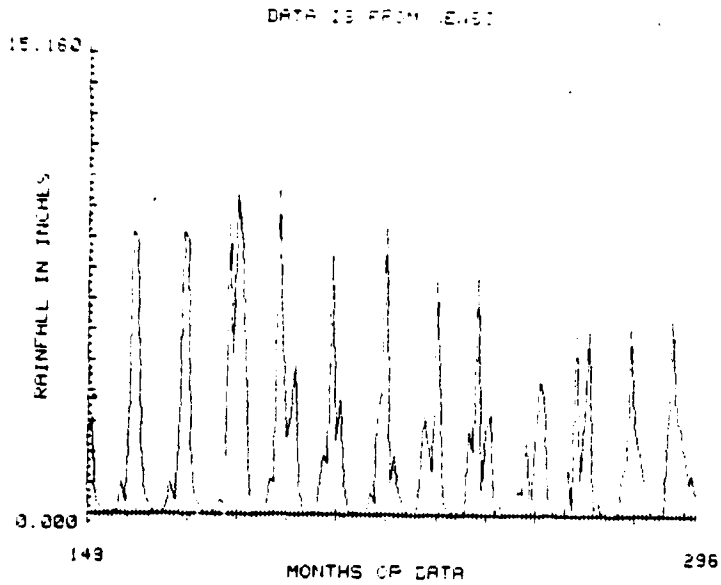
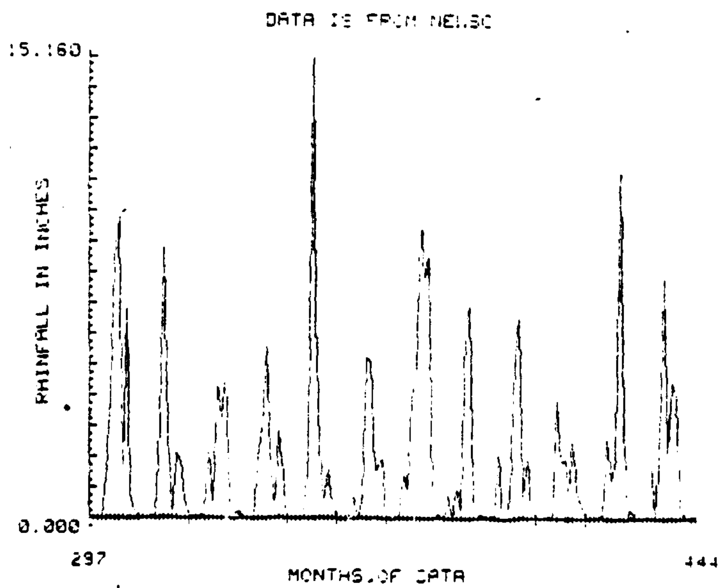


Figure 25a. Months 1 (October 1926) - 148 (January 1938) of rainfall in inches for data set SC.





Fiugre 25b. Months 149 - 444 of rainfall for data set SC

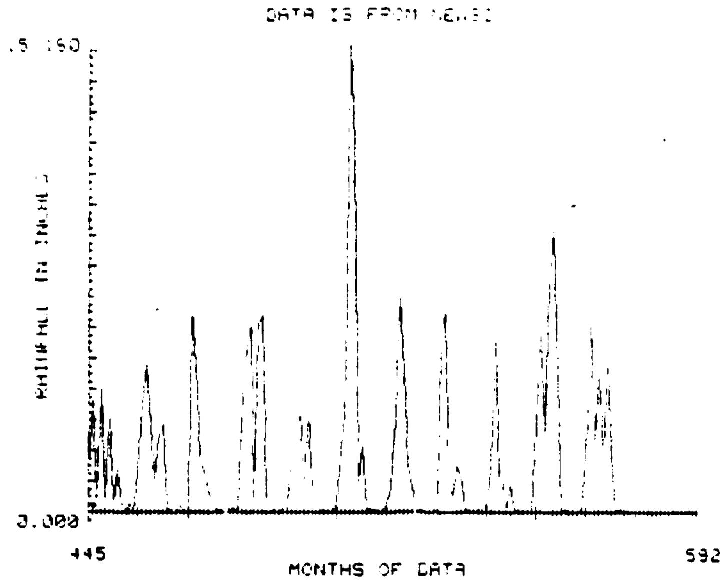


Figure 25c. Months 445 - 576 of rainfall for data set SC

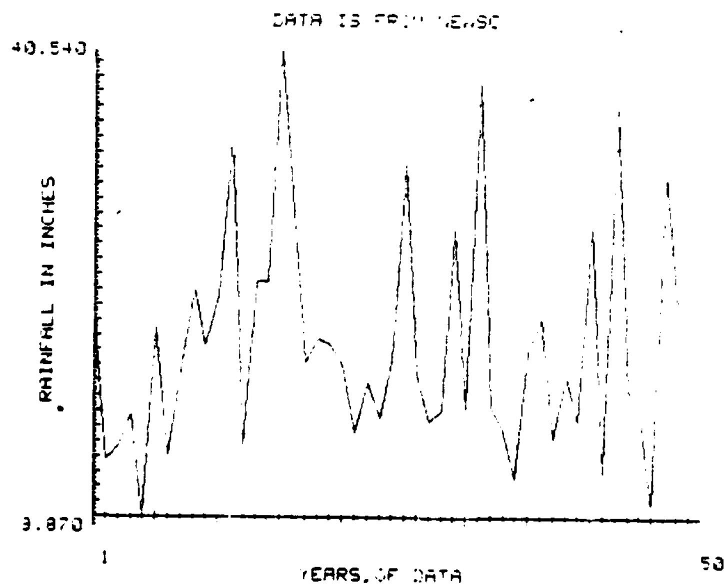


Figure 26. Yearly total rainfall for data set SC (1926 - 1974)

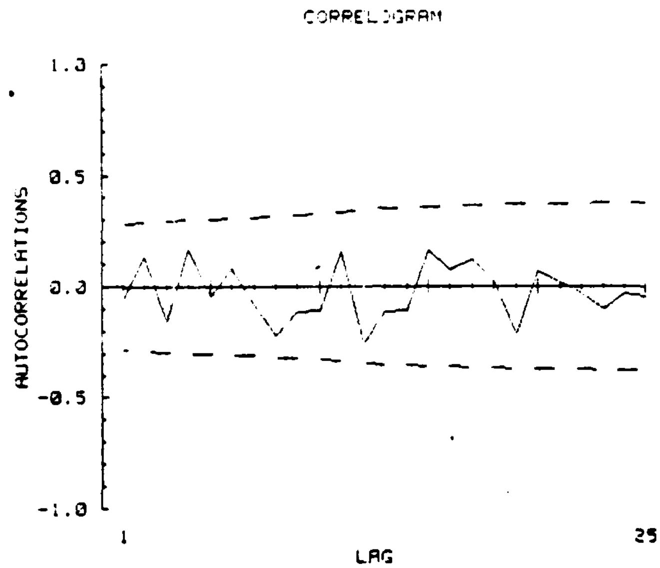


Figure 27. Correlogram of yearly total rainfall for data set SC

TABLE 11

ESTIMATED AUTOCORRELATIONS OF YEARLY TOTAL RAINFALL FOR DATA SET SC

AUTOCORRELATIONS

LAG	VALUE	LAG	VALUE
1	-.050	14	-.109
2	.135	15	.161
3	-.158	16	.077
4	.168	17	.116
5	-.042	18	.025
6	.081	19	-.214
7	-.084	20	.066
8	-.217	21	.019
9	-.111	22	-.028
10	-.107	23	-.101
11	.158	24	-.030
12	-.260	25	-.042
13	-.110		

2. Swept Data

TABLE 12

MONTHLY MEANS AND VARIANCES
FOR DATA SET SC

MONTH	MEAN	VARIANCE
1	.698	.5945
2	2.175	4.2382
3	3.940	10.1783
4	4.599	8.4899
5	4.353	13.3443
6	3.080	5.2744
7	1.700	3.4486
8	.431	.1912
9	.111	.0451
10	.017	.0055
11	.037	.0125
12	.103	.1040

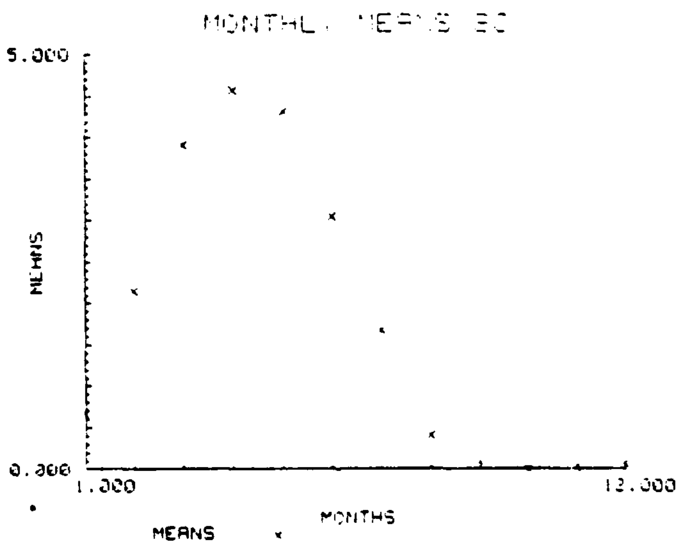


Figure 28. Monthly means for data set SC

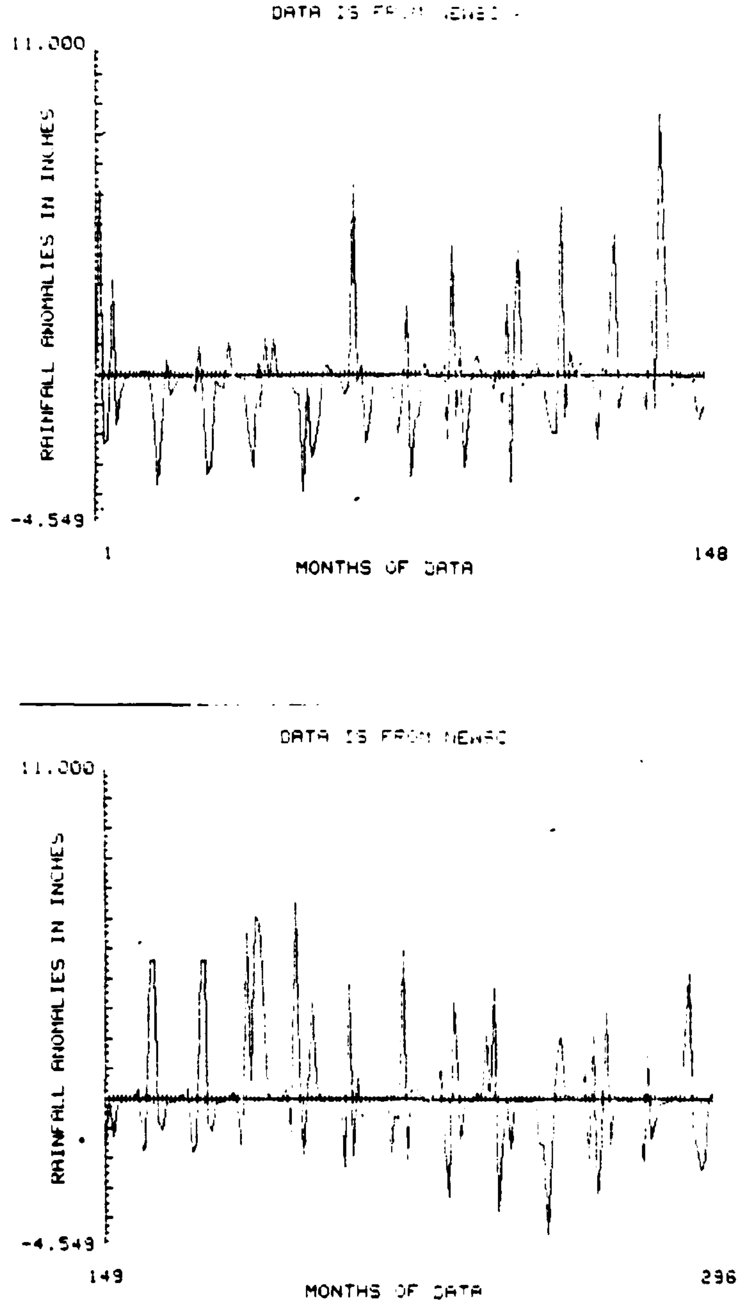
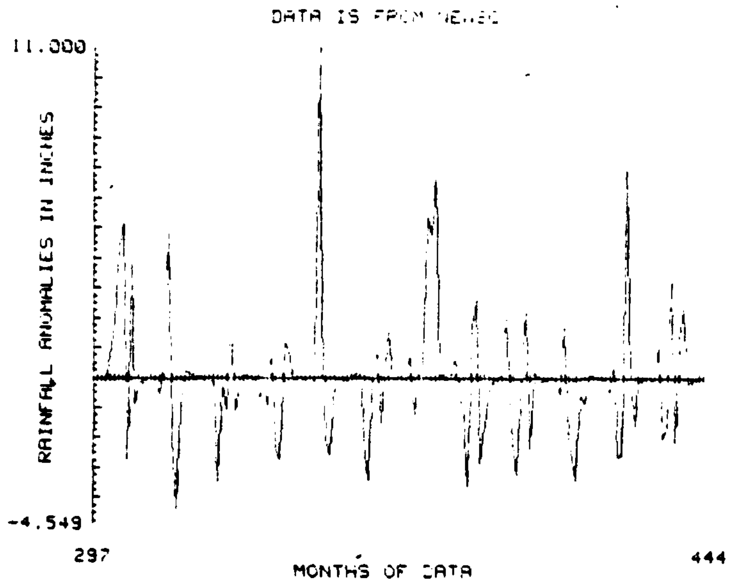


Figure 29a. Months 1 - 296 anomalies in inches for data set SC



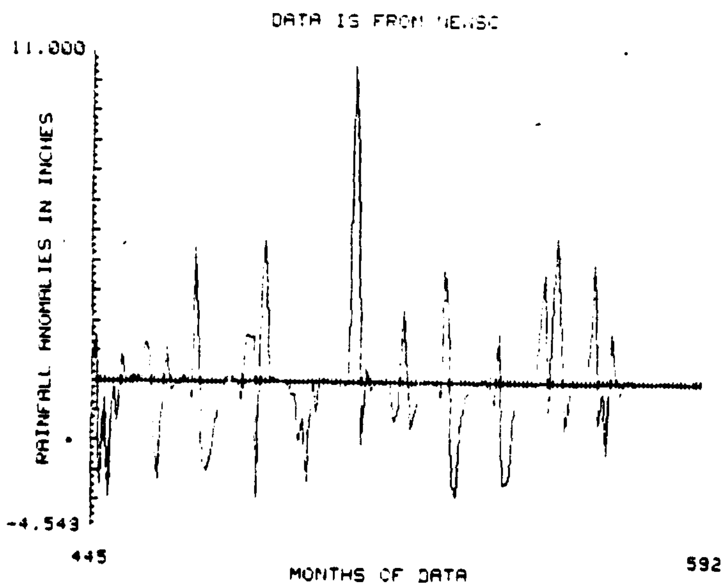


Figure 29b. Months 297 - 576 anomalies in inches for data set SC

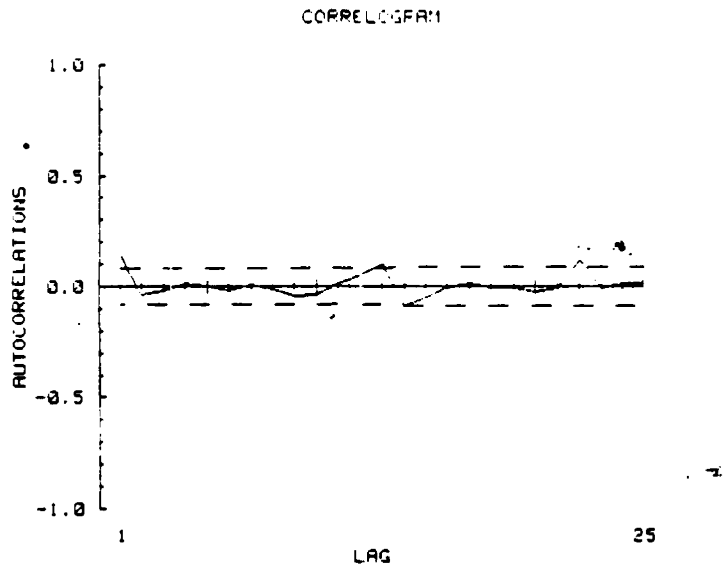


Figure 30. Correlogram of monthly rainfall anomalies for data set SC

TABLE 13

ESTIMATED AUTOCORRELATIONS OF MONTHLY RAINFALL ANOMALIES FOR DATA SET SC

AUTOCORRELATIONS

LAG	VALUE	LAG	VALUE
1	.140	14	-.088
2	-.039	15	-.051
3	-.021	16	-.006
4	.012	17	.015
5	-.001	18	-.006
6	-.019	19	-.008
7	.003	20	-.027
8	-.013	21	-.003
9	-.038	22	-.122
10	-.038	23	-.006
11	.014	24	.011
12	.051	25	.011
13	.102		

3. Logged and Swept Data

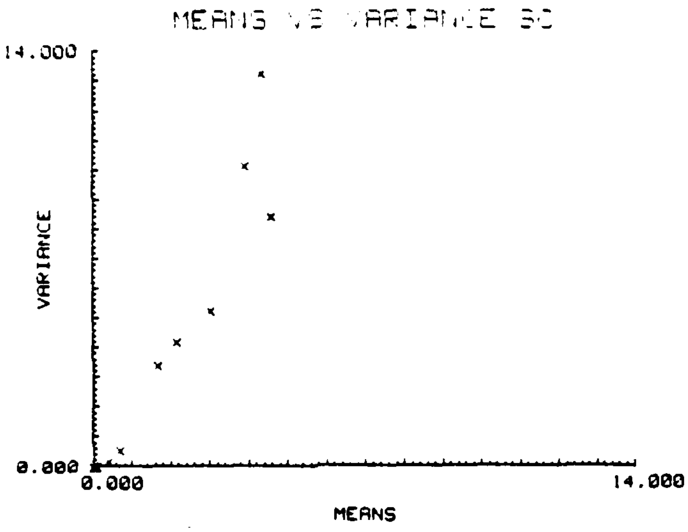


Figure 31. Plot of monthly variance against anomally means for data set SC.

TABLE 14

MONTHLY MEANS AND VARIANCES OF LOGGED DATA SET SC

MONTH	MEAN	VARIANCE
1	.444	.1623
2	.961	.3999
3	1.408	.3949
4	1.583	.3117
5	1.444	.4928
6	1.243	.3556
7	.817	.3247
8	.320	.0726
9	.092	.0227
10	.015	.0040
11	.031	.0083
12	.075	.0351

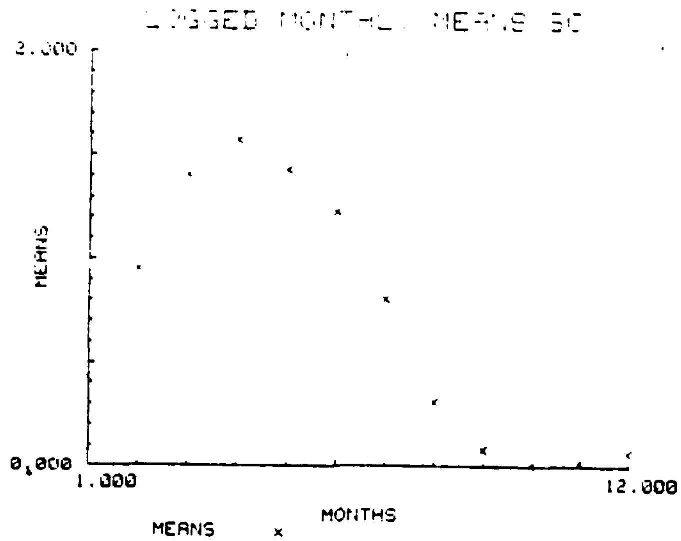


Figure 32. Monthly means of logged data set SC

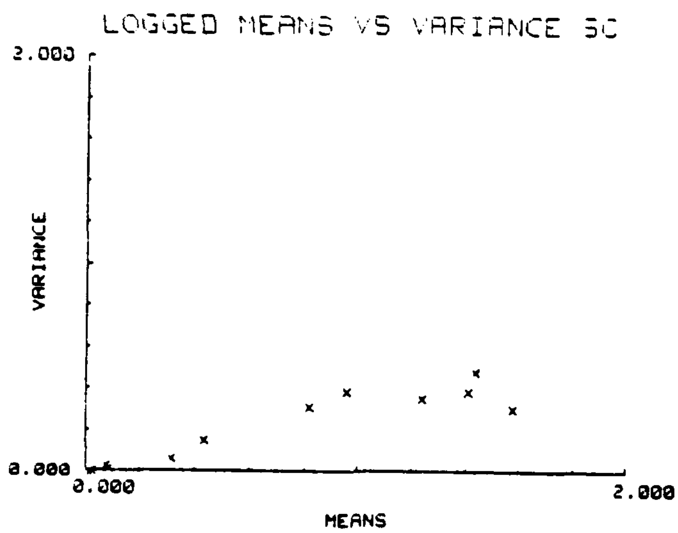


Figure 33. Plot of monthly variance against monthly means for logged data set SC

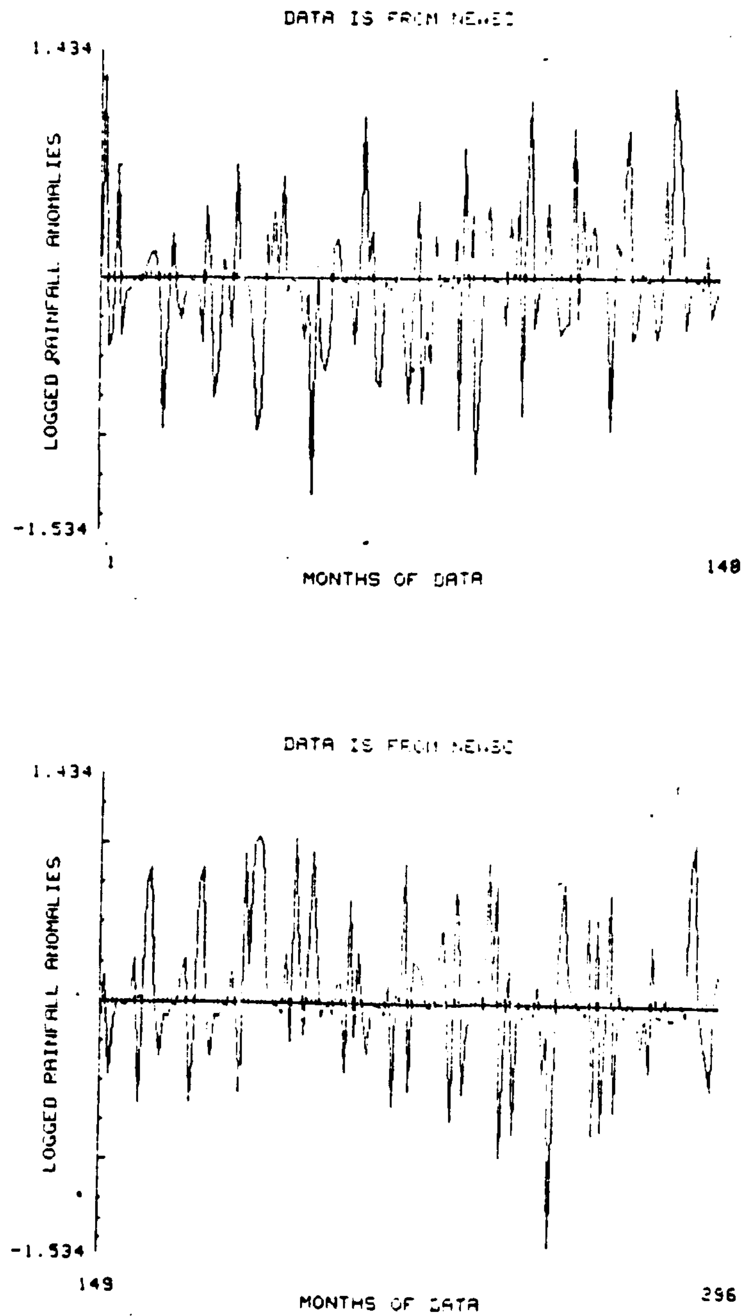
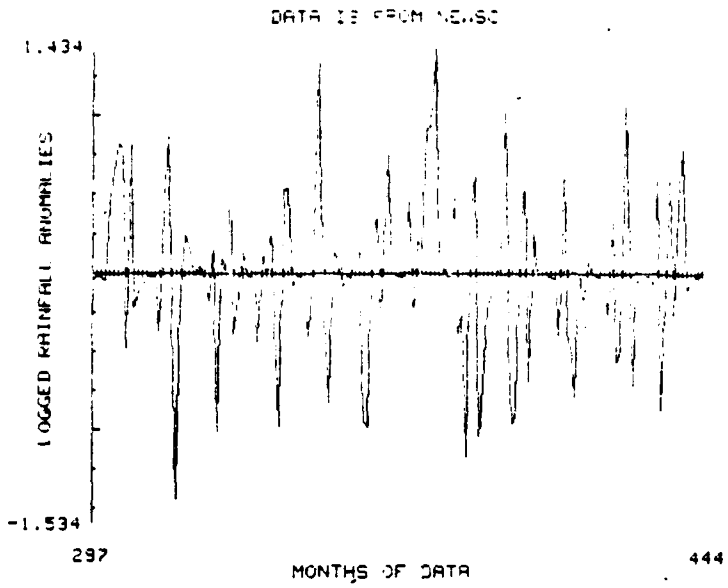


Figure 34a. Months 1 - 296 of logged rainfall anomalies from data set SC



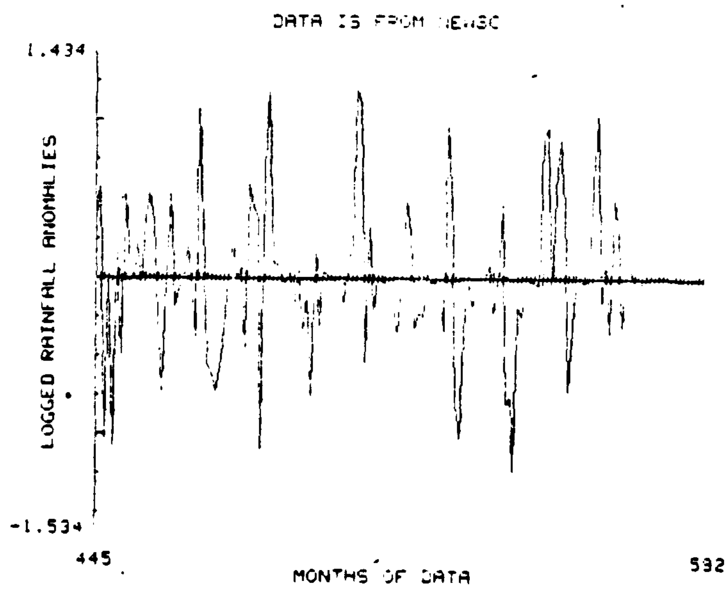


Figure 34b. Months 297 - 576 of logged rainfall anomalies from data set SC

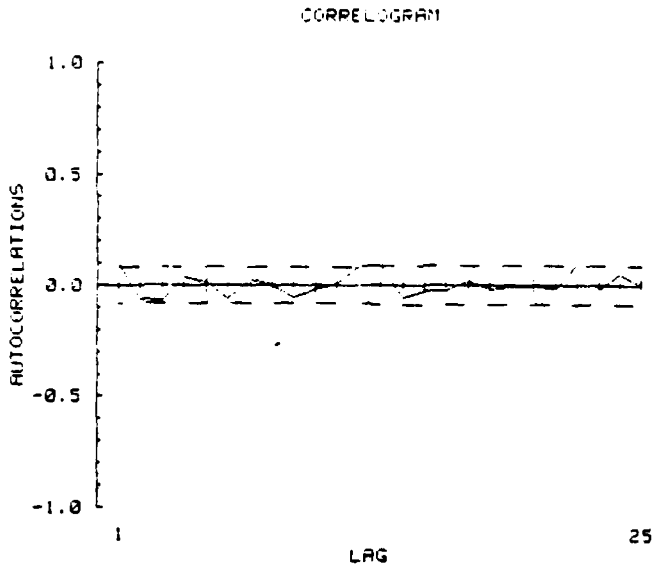


Figure 35. Correlogram of logged anomalies of monthly rainfall from data set SC

TABLE 15

ESTIMATED AUTOCORRELATION OF LOGGED ANOMALIES OF MONTHLY RAINFALL FROM DATA SET SC

AUTOCORRELATIONS

LAG	VALUE	LAG	VALUE
1	.096	14	-.057
2	-.065	15	-.022
3	-.066	16	-.023
4	.038	17	.021
5	.012	18	-.017
6	-.061	19	-.005
7	.023	20	-.008
8	-.001	21	-.011
9	-.056	22	.092
10	-.021	23	-.019
11	.007	24	.050
12	.091	25	.002
13	.091		

III. FIRST ORDER MARKOV MODEL

A. THEORY

As first shown by equation I.1, the general ARMA(p,q) model is:

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \ldots + \phi_p \tilde{z}_{t-p} + a_t - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q} \qquad \text{III.1}$$

The development and discussion of this type of model is contained in detail in Box and Jenkin [Ref. 1] and Nelson [Ref. 8]. The modeling process is a three fold procedure. The parts are:

(1) Identification

(2) Estimation

(3) Diagnosis.

Identification is conducted using the correlogram and a plot of the partial-autocorrelations (or partial correlogram). The partial autocorrelations are related to the autocorrelations, see Box and Jenkins [Ref. 1], Nelson [Ref. 8], or Richards and Woodall [Ref. 12]. These partial autocorrelations are used to determine the order of the moving average process much like the autocorrelations may be used to determine the order of the auto-regressive process.

Once the autocorrelations and partial autocorrelations have been found, the degree of the ARMA may be estimated by

techniques described in Box and Jenkins, Nelson or Richards and Woodall. Each of the data sets, once logged and swept, indicated that the most probable model was an ARMA(1,0) or AR(1) or more commonly a first-order autoregressive Markov model. This model is simply;

$$\tilde{z}_t = \rho\tilde{z}_{t-1} + a_t \qquad \text{III.2}$$

where the ρ is the autocorrelation of lag one. Thus, this model indicates that any persistence in the data are conditionally independent of the past given the lag one value.

Subsections B, C, and D below show this model as applied to the three data sets of interest. The residuals of the model $\tilde{z}_t - \rho\tilde{z}_{t-1}$ are examined. The residuals appear to be independent, however, they do not appear to be normally distributed; for example, there is a high peak around zero. One possible reason for this discrepancy may be the dichotomy of winter and summer rain as indicated in Tables 2, 4, 7, 9, 12, and 14. The existence of months with zero rainfall during the summer suggests that one should consider the summer, when rain is sparse, completely separate from the winter when rain is more abundant. Therefore, also shown in the subsections below is the autoregressive model applied to the winter months only. This is accomplished by stripping out months 9 through 12 (June through September) of the data sets and treating the remaining data as a continuous set. In other words, the first ten months are then

$R_{1,1}, R_{1,2}, R_{1,3}, R_{1,4}, R_{1,5}, R_{1,6}, R_{1,8}, R_{2,1}, R_{2,2}.$

The appropriate correlograms and partial correlograms are displayed prior to the model applications.

B. DATA SET RN

1. Twelve Month Series

This data set is described in section II.b. The remaining diagnostic device needed is the partial correlogram of Figure 36 and the corresponding values in Table 16.

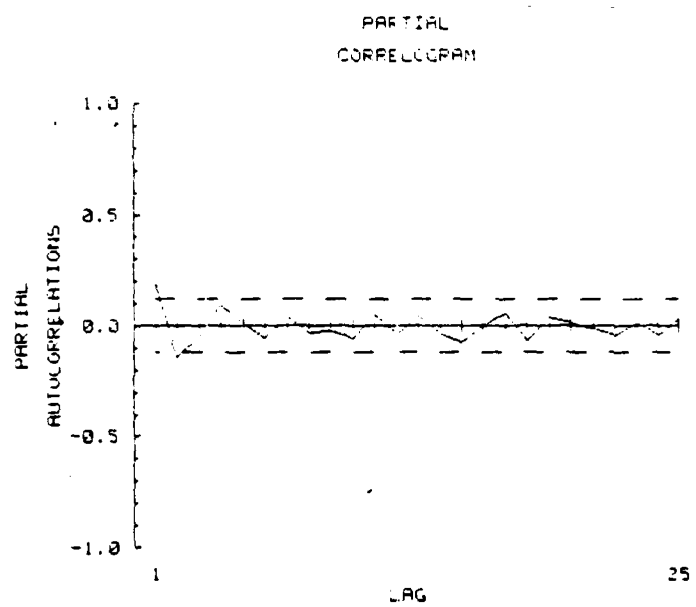


Figure 36. Partial correlogram of the logged rainfall anomalies of data set RN

TABLE 16

ESTIMATED PARTIAL-AUTOCORRELATIONS FOR
LOGGED RAINFALL ANOMALIES OF DATA SET RN

LAG	VALUE	LAG	VALUE
1	.191	14	-.040
2	-.144	15	-.073
3	-.047	16	.001
4	.092	17	.054
5	.006	18	-.067
6	-.058	19	.039
7	.037	20	.012
8	-.034	21	-.001
9	-.028	22	-.043
10	-.056	23	.013
11	.048	24	-.043
12	-.041	25	-.034
13	.047		

The model of interest is then

$$\tilde{Z}'_t = .191\tilde{Z}'_{t-1} + a_t \qquad \text{III.3}$$

where the random shocks $\{a_t\}$ are assumed to be distributed iid $N(0,\sigma_a^2)$ and σ_a^2 is estimated as

$$\frac{1}{N-1}\sum_{t=1}^{N}(\tilde{Z}'_t - .191\tilde{Z}'_{t-1}) \qquad \text{III.4}$$

The goodness of this fit may be viewed in two ways. Firstly, are the residuals, $\{a_t\}$ independent? Secondly, are the residuals distributed as Normal (Gaussian) random variables? A plot of the residuals follows in Figure 37. The question of independence is addressed in Figure 38 (Correlogram), Figure 39 (Lag one plot), Figure 40 (Residuals vs. lag one), and Table 17 (Turning points). For a discussion of the usefulness of the turning points see Kendall [Ref. 14].

All of these displays and tests tend to indicate that the residuals are in fact serially independent. The statistics of the residuals are in Table 18. A Normal Plot of the residuals (Figure 41), in which the sample is normalized by removing the mean and scaling by the standard deviation and then plotted on normal paper, should yield a nearly straight line corresponding to the dashed line of the figure. The Normal Plot accompanied by the sample histogram (Figure 42) addresses the normality of these data. As may be seen from the kurtosis, the fluctuations of the sample CDF near the midpoint, and the peak of the histogram, the normality of this data are questionable. To confirm this a chi-squared goodness of fit test was conducted yielding a value of 49.18 with 17 degrees of freedom, again rejecting any hypothesis of normality at a significance level of $5x10^{-5}$.

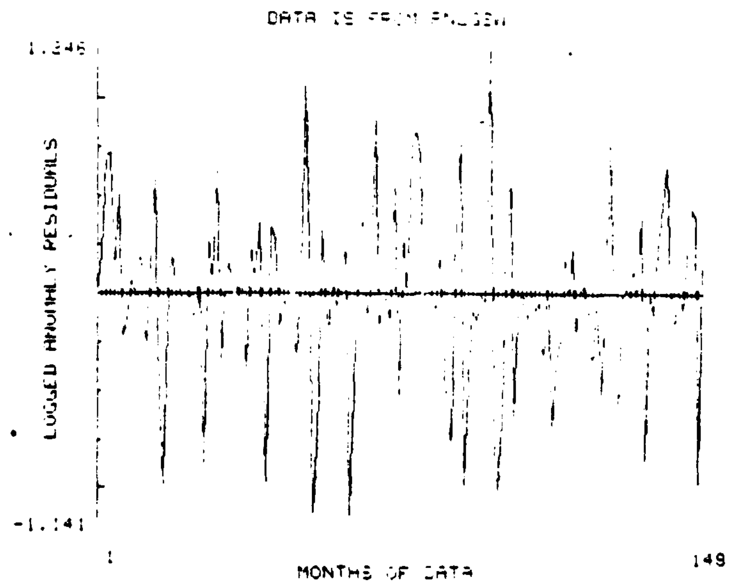


Figure 37. First order Markov residuals from logged rainfall anomalies of data set RN. Months 149 - 292

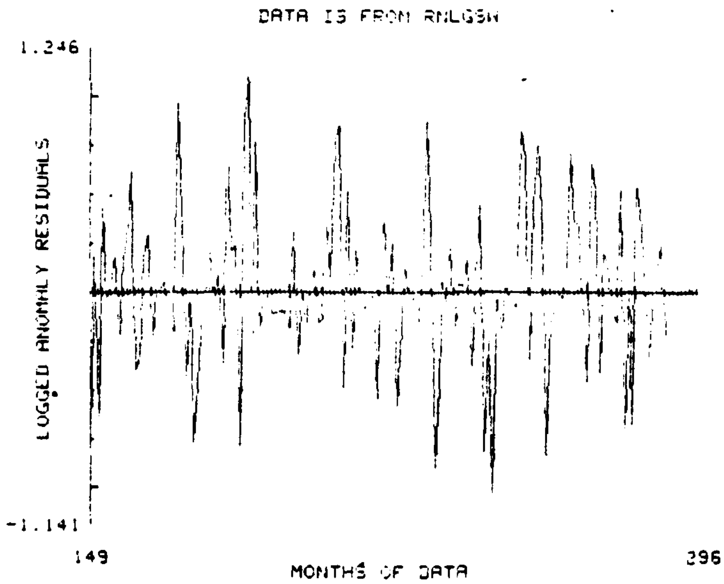


Figure 37b. First order Markov residuals from logged rainfall anomalies of data set RN. Months 149 - 292

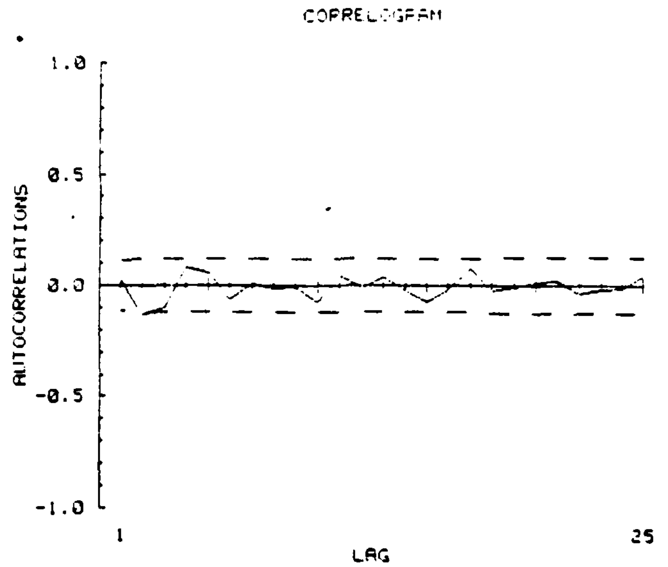


Figure 38. Auto correlations of residuals from first order Markov process applied to the logged rainfall anomalies of data set RN

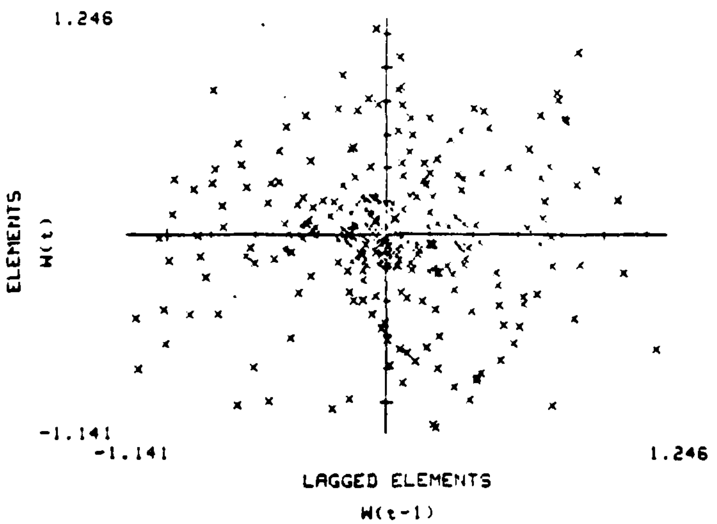


Figure 39. Lag one plot of first order Markov residuals from logged rainfall anomalies of data set RN

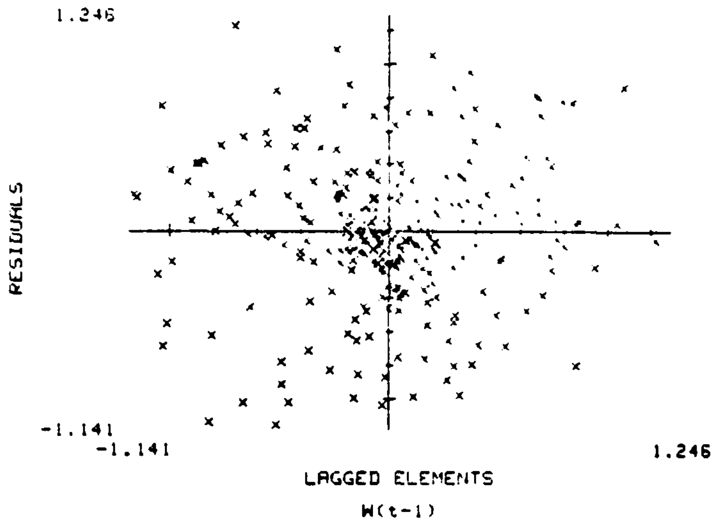


Figure 40. First order Markov residuals versus lag one data point from logged rainfall anomalies of data set RN

TABLE 17

ACTUAL AND EXPECTED NUMBER OF TURNING POINTS AND ACTUAL AND EXPECTED PHASE FREQUENCIES FOR THE FIRST ORDER MARKOV RESIDUALS FROM THE LOGGED RAINFALL ANOMALIES OF DATA SET RN

NUMBER OF TURNING POINTS = 191
E[P] = 190.667 V[P] = 15.899

PHASE LENGTHS

D	OBS.	E[*]
1	117	118.8
2	56	52.1
3	15	14.9
4	3	3.2
5	0	.6
6	0	.1
7	0	0.0

TABLE 18

GENERAL STATISTICS OF FIRST ORDER MARKOV RESIDUALS FROM LOGGED RAINFALL ANOMALIES OF DATA SET RN

Moments

Mean	-.001
Variance	.117
Skewness	-.066
Kurtosis	.523

Percentiles

Minimum	-1.141
Lower Sixteenth	- .745
Lower Eight	- .463
Lower Quartile	- .174
Median	- .014
Upper Quartile	.211
Upper Eight	.436
Upper Sixteenth	.706
Maximum	1.246

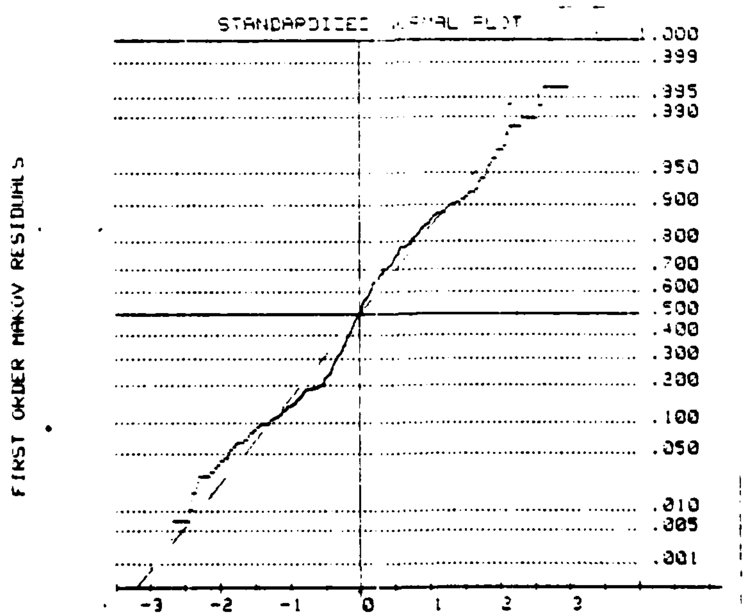


Figure 41. Standardized normal plot of first order Markov residuals from logged rainfall anamlies of data set RN

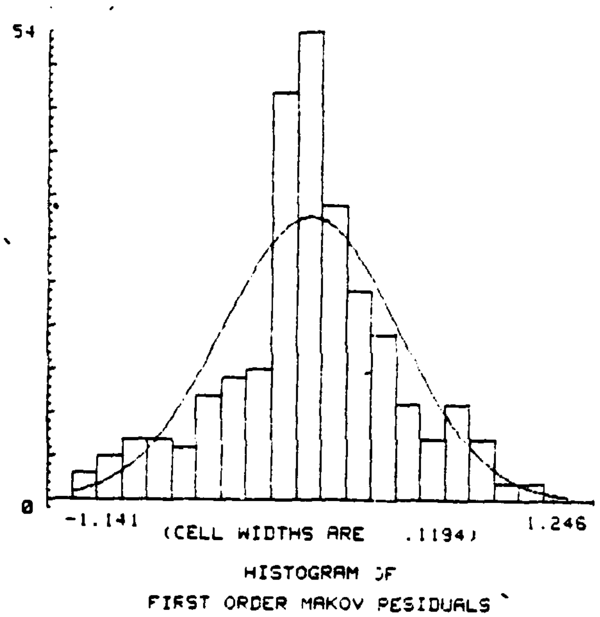


Figure 42. Histogram of first order Markov residuals from logged rainfall anomalies of data set RN

2. <u>Winter Series</u>

As stated above, the number of summer months with zeros indicated that a look at the winter months only might be worthwhile. The Figures 43 (Winter months), 44 (Correlogram), and 45 (Partial correlogram), which deal only with winter months, still indicate a first order autoregressive Markov model as;

$$\tilde{Z}''_t = .218\tilde{Z}''_{t-1} + a_{t-1} \qquad \text{III.5}$$

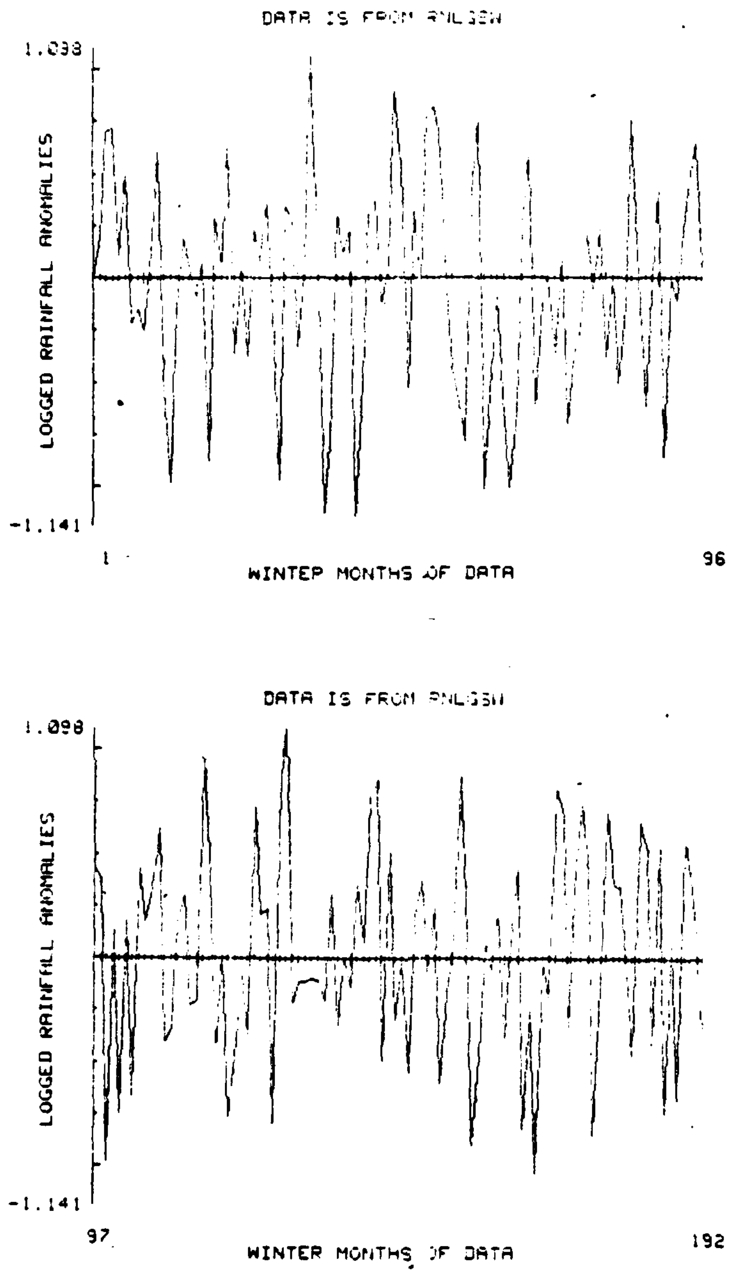


Figure 43. Winter months only of logged rainfall anomalies of data set RN

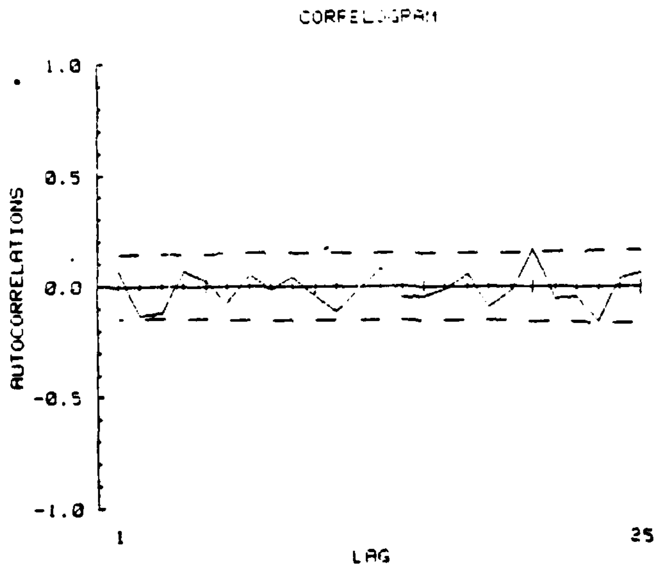


Figure 44. Correlogram of winter months only of logged rainfall anomalies from data set RN

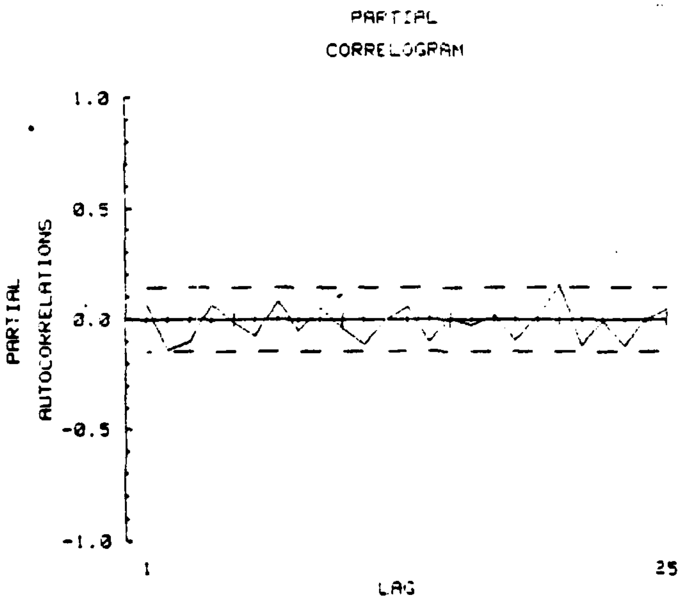


Figure 45. Partial autocorrelations of winter months only logged rainfall anomalies from data set RN

Now, as with the full twelve month model, a look at the residuals yields the Figures 46 (Residuals), 47 (Correlogram), and 48 (Lag one plot), 49 (Residuals vs. lag one), and Table 19 (Turning points). It appears that the residuals are, in fact, independent. This is similar to the twelve month model.

The question of the normality of the residuals is addressed by Table 19 and Figures 50 (Normal plot) and 51 (Histogram). The results of these plots and a basic chi-squared goodness of fit of 22.11 with 17 degrees of freedom indicate that this winter month data set is much more normal than was its twelve month counterpart. This chi-squared value is significant at the .181 level.

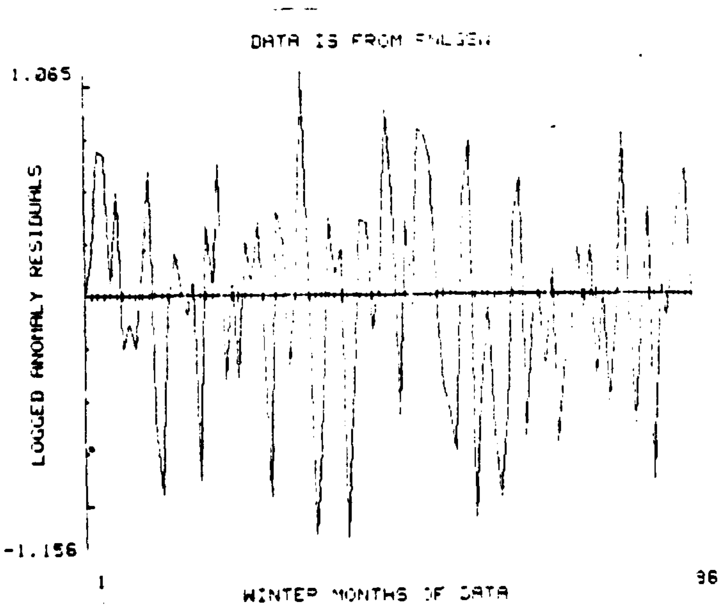


Figure 46a. First order Markov residuals of logged rainfall anomalies for winter months only of data set RN. Years 1 - 12

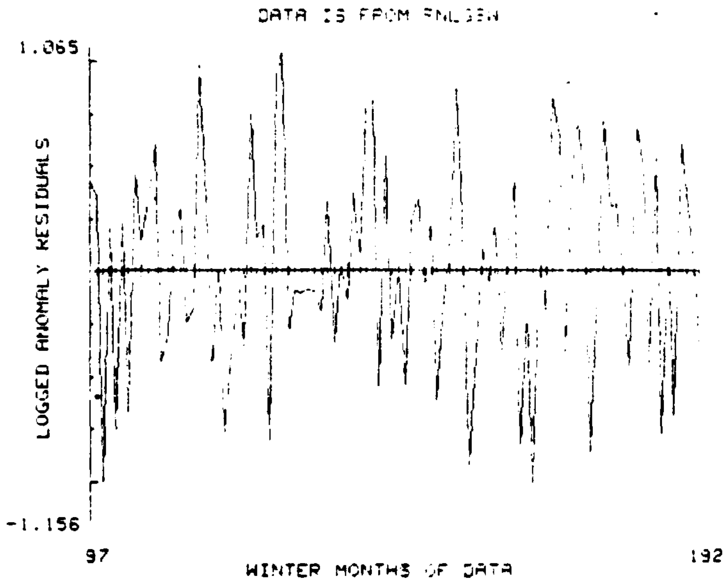


Figure 46b. First order Markov residuals of logged rainfall anomalies for winter months only of data set RN. Years 12 - 24

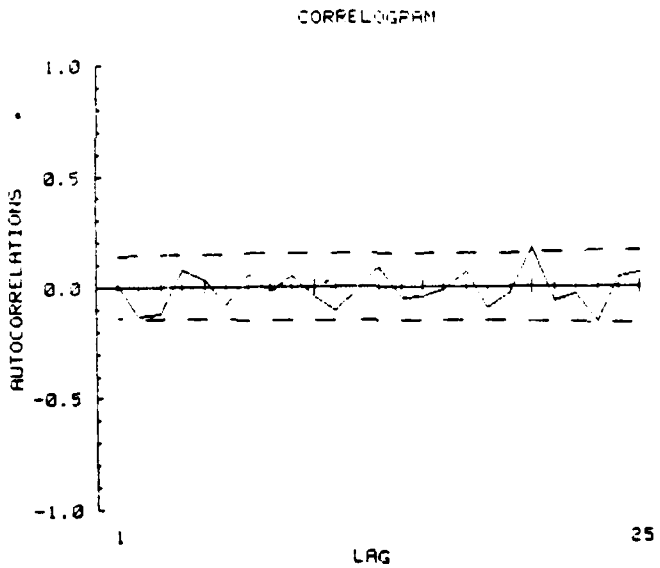


Figure 47. Correlogram of first order Markov residuals of logged rainfall anomalies for winter months only of data set RN

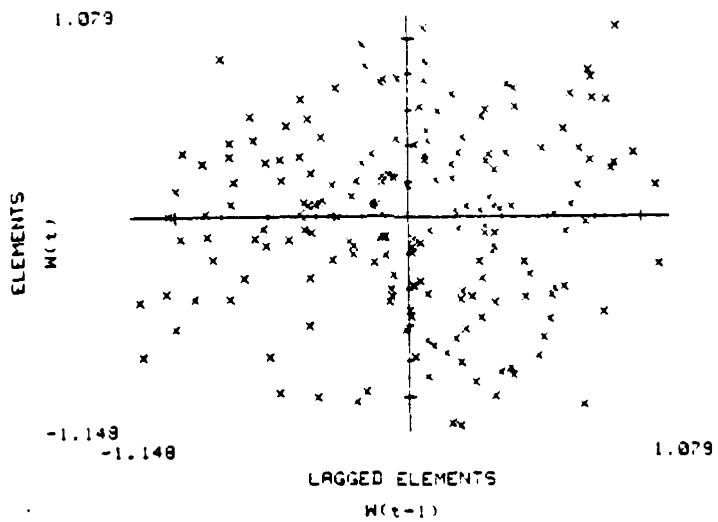


Figure 48. Lag one plot of first order Markov residuals from logged rainfall anomalies for winter months only of data set RN

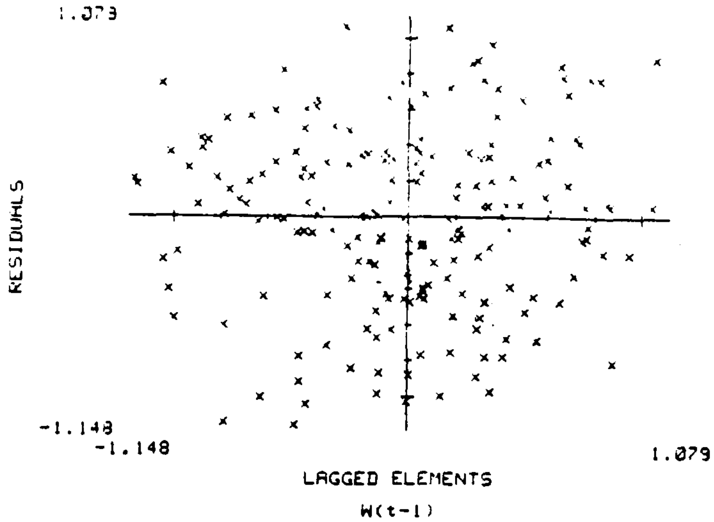


Figure 49. First order Markov residuals versus lag one data point from lagged rainfall anomalies of winter month only data from data set RN

TABLE 19

ACTUAL AND EXPECTED NUMBER OF TURNING POINTS AND ACTUAL AND EXPECTED PHASE FREQUENCIES FROM THE FIRST ORDER MARKOV RESIDUALS OF THE LOGGED RAINFALL ANOMALIES OF DATA SET RN

NUMBER OF TURNING POINTS = 129

E[P] = 126.667   V[P] = 15.84

PHASE LENGTHS

D	OBS.	E[*]
1	82	78.8
2	38	34.5
3	7	9.9
4	1	2.1
5	1	.4
6	0	0.0
7	0	0.0

TABLE 20

GENERAL STATISTICS OF FIRST ORDER MARKOV RESIDUALS FROM LOGGED RAINFALL ANOMALIES OF DATA SET RN

Moments

Mean	-.001
Variance	.247
Skewness	-.161
Kurtosis	-.521

Percentiles

Maximum	-1.148
Lower Sixteenth	- .845
Lower Eight	- .616
Lower Qurtile	- .368
Median	.025
Upper Quartile	.339
Upper Eight	.591
Upper Sixteenth	.762
Maximum	1.078

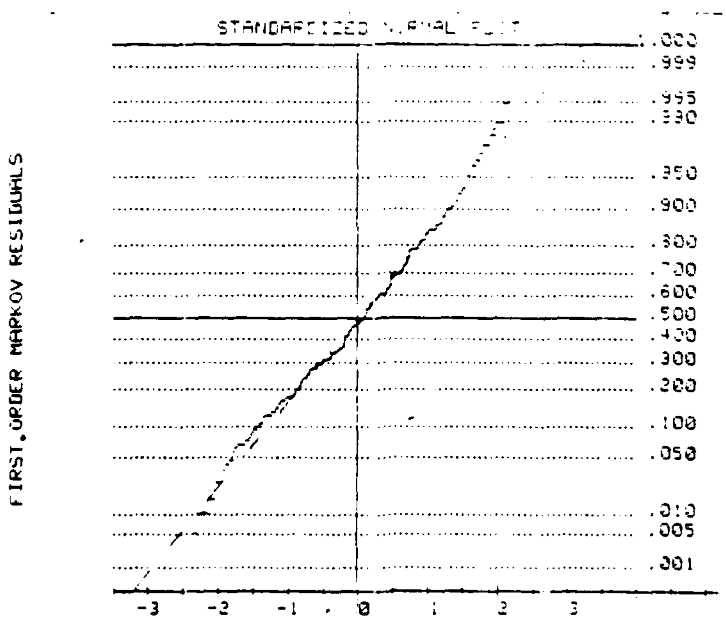


Figure 50. Standardized normal plot of first order Markov residuals from logged rainfall anomalies of winter months only from data set RN

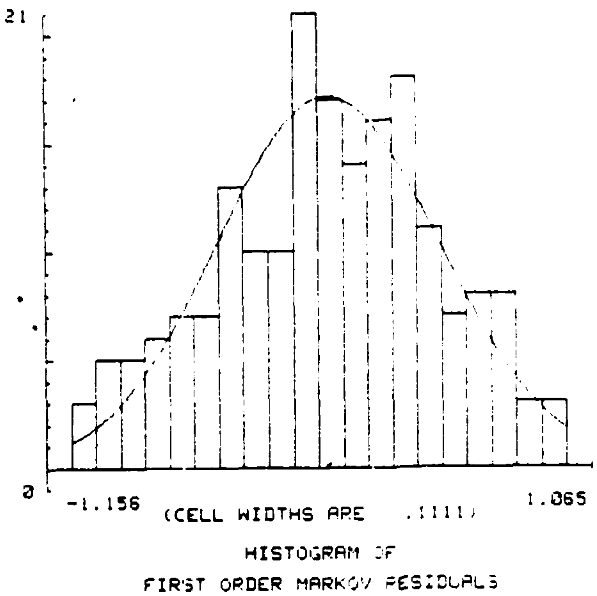


Figure 51. Histogram of first order Markov residuals from logged rainfall anomalies of winter month only of data set RN

C. DATA SET FL

As in the previous section on the data sets, the analysis of section III.B above carries forward fairly well to data sets FL and SC. This section, and the following, contain only the Figures and Tables corresponding to those in the previous section on data set RN.

1. <u>Twelve Month Series</u>

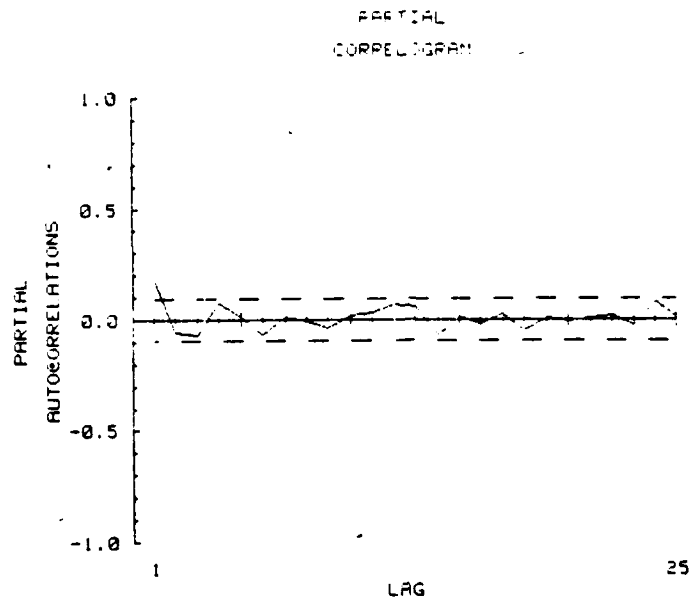


Figure 52. Partial correlogram of the logged rainfall anomalies for data set FL

TABLE 21

ESTIMATED PARTIAL AUTOCORRELATIONS FOR LOGGED RAINFALL ANOMALIES OF DATA SET FL

LAG	VALUE	LAG	VALUE
1	.185	14	-.072
2	-.057	15	.011
3	-.069	16	-.019
4	.077	17	.027
5	.016	18	-.048
6	-.068	19	.006
7	.010	20	-.007
8	-.008	21	.003
9	-.040	22	.018
10	.022	23	-.025
11	.031	24	.081
12	.067	25	.012
13	.055		

$$\tilde{Z}_t' = .185\tilde{Z}_{t-1}' + a_t \qquad \text{III.6}$$

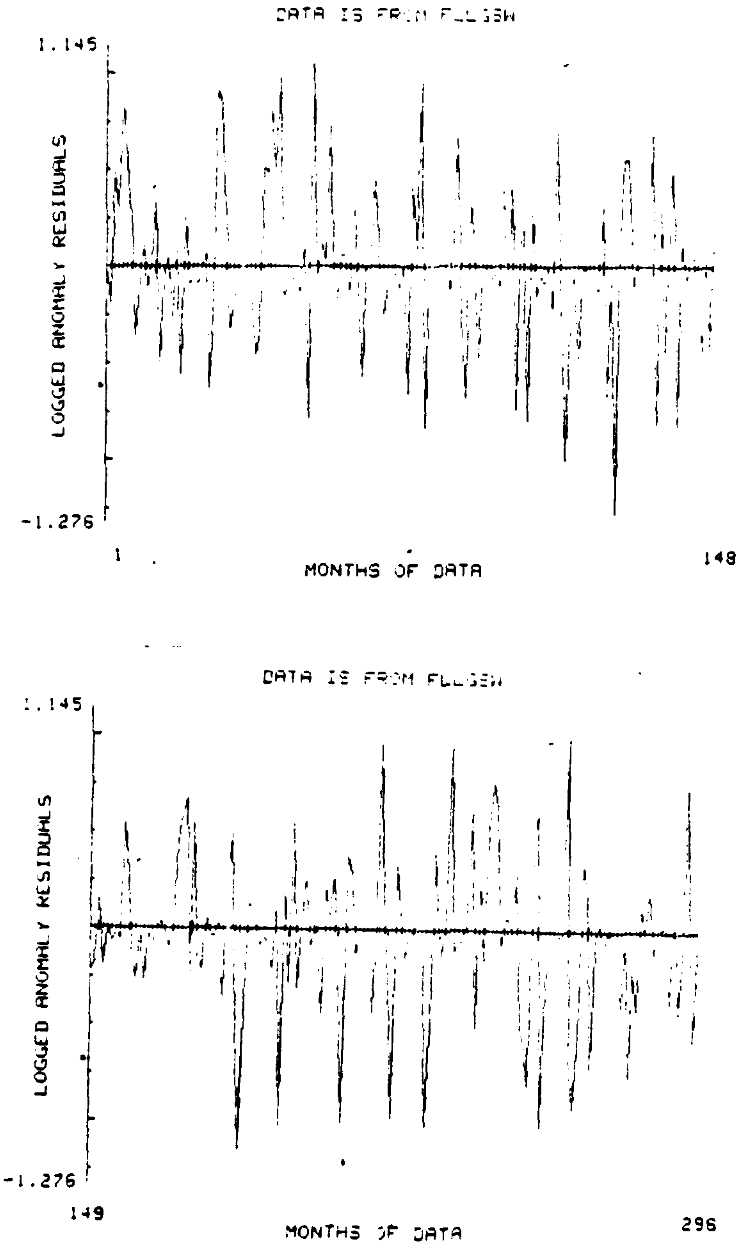


Figure 53a. First order Markov residuals from logged rainfall anomalies of data set FL. Months 1 - 296

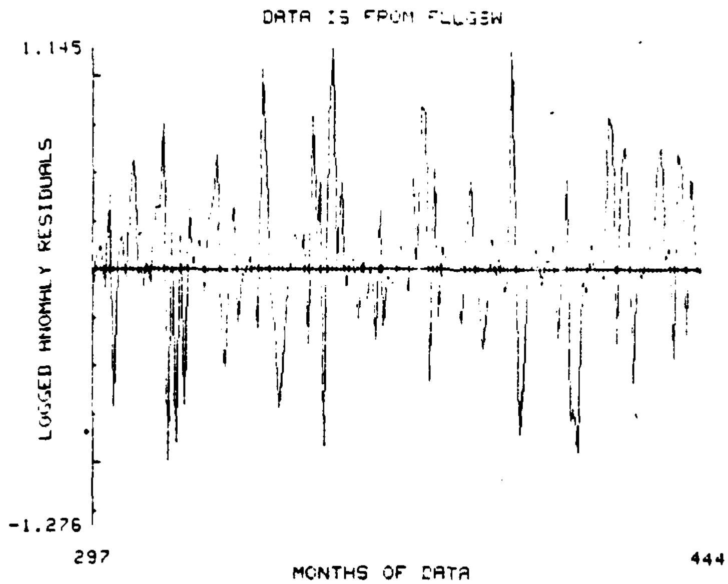


Figure 53b. First order Markov residuals from logged rainfall anomalies of data set FL. Months 297 - 444.

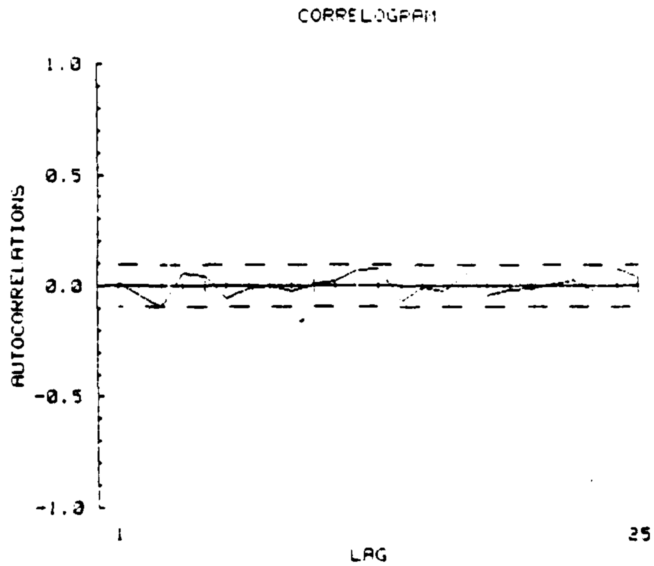


Figure 54. Autocorrelations of residuals from first order Markov process applied to the logged rainfall anomalies of data set FL

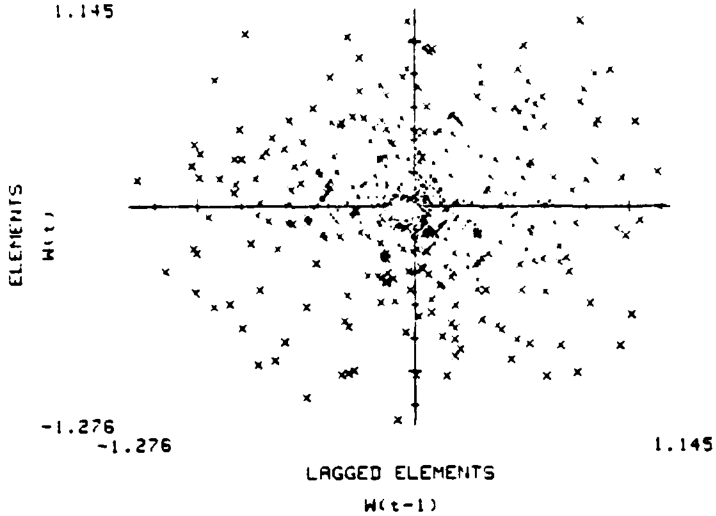


Figure 55. Lag one plot of first order Markov residuals from logged rainfall anomalies of data set FL

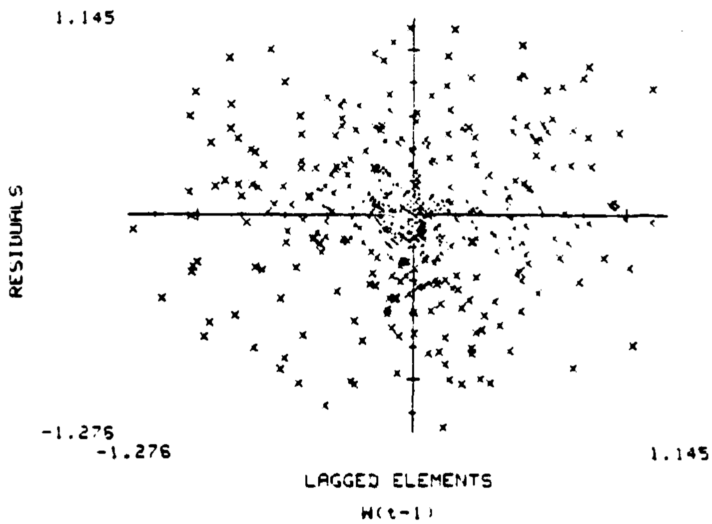


Figure 56. First order Markov residuals versus lag one data points from logged rainfall anomalies of data set FL

TABLE 22

ACTUAL AND EXPECTED NUMBER OF TURNING POINTS AND ACTUAL AND EXPECTED PHASE FREQUENCIES FROM THE FIRST ORDER MARKOV RESIDUALS FROM DATA SET FL

NUMBER OF TURNING POINTS = 294
E[P] = 294.667   V[P] = 15.93

PHASE LENGTHS

D	OBS.	E[*]
1	188	183.8
2	79	80.7
3	19	23.2
4	6	5.0
5	0	.9
6	1	.1
7	1	0.0
8	0	0.0
9	0	0.0
10	0	0.0
TOTALS	294	293.7

TABLE 23

GENERAL STATISTICS OF FIRST ORDER MARKOV RESIDUALS FROM LOGGED RAINFALL ANOMALIES OF DATA SET FL

Moments

Mean	.000
Variance	.163
Skewness	-.044
Kurtosis	.717

Percentiles

Minimum	-1.276
Lower Sixteenth	- .702
Lower Eight	- .426
Lower Quartile	- .156
Median	- .012
Upper Quartile	.184
Upper Eight	.481
Upper Sixteenth	.648
Maximum	1.124

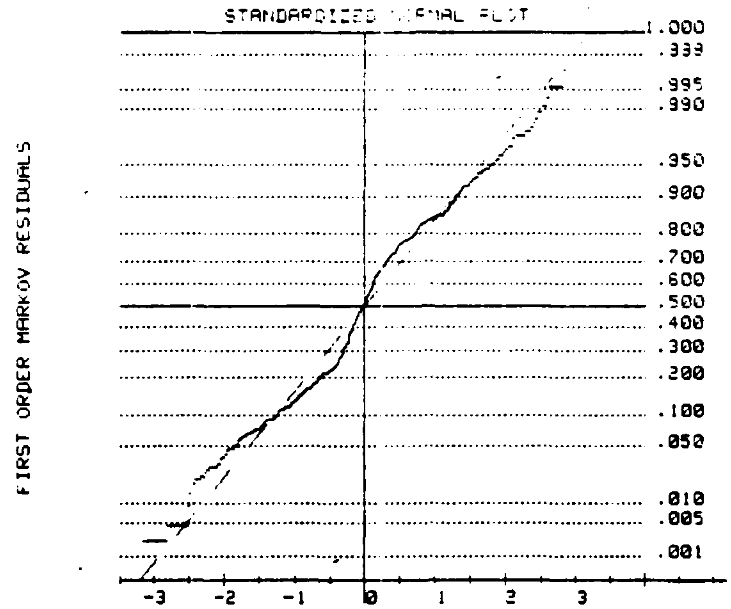


Figure 57. Standardized normal plot of first order Markov residuals from logged rainfall anomalies of data set FL

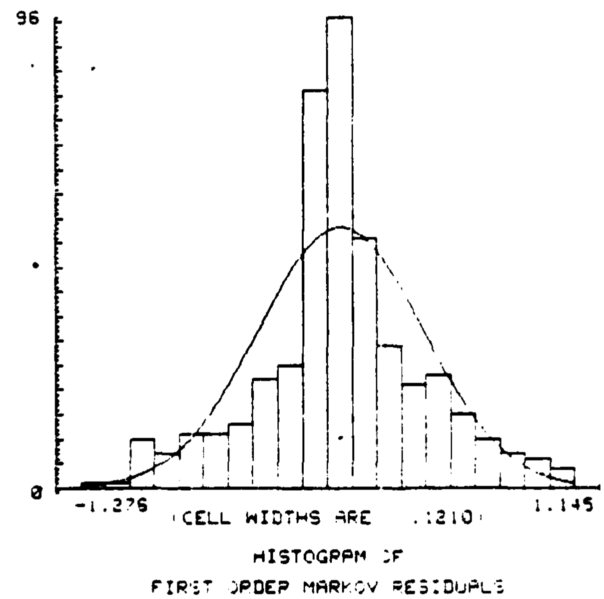


Figure 58. Histogram of first order Markov residuals from logged rainfall anomalies of data set FL

This data set yielded a chi-square value of 107.66 for 17 degrees of freedom. The significance of the value is in the zero plus range.

2. Winter Series

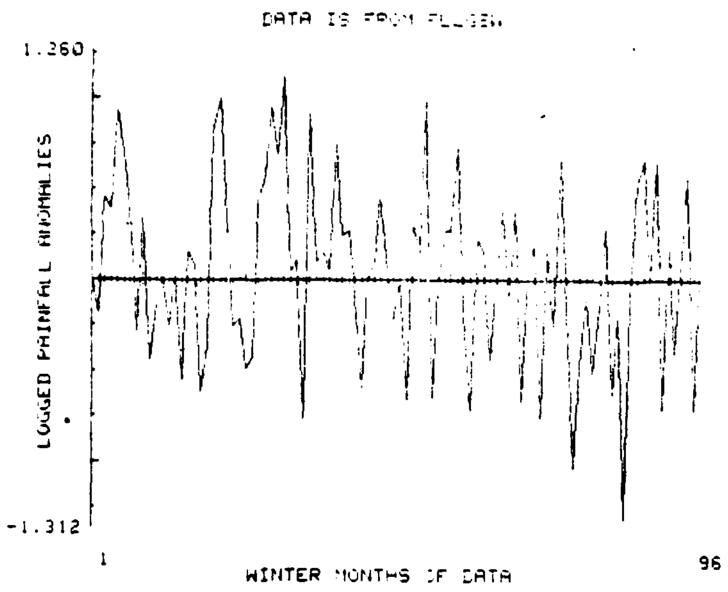
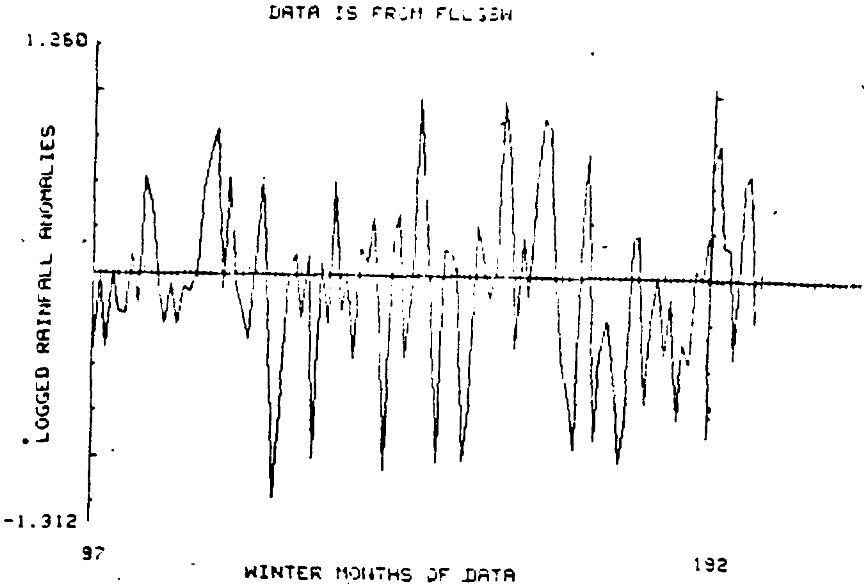


Figure 59a. Years 1 - 12 of winter months only of logged rainfall anomalies of data set FL



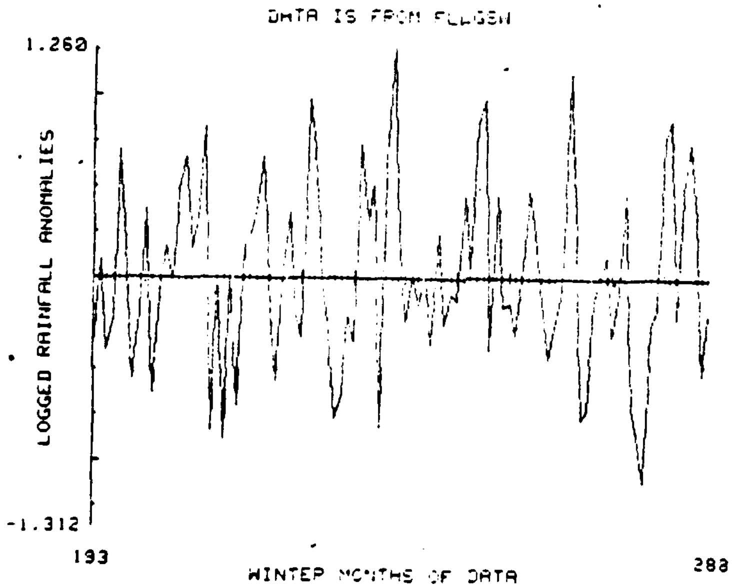


Figure 59b. Years 13 - 37 of winter months only, logged rainfall anomalies of data set FL

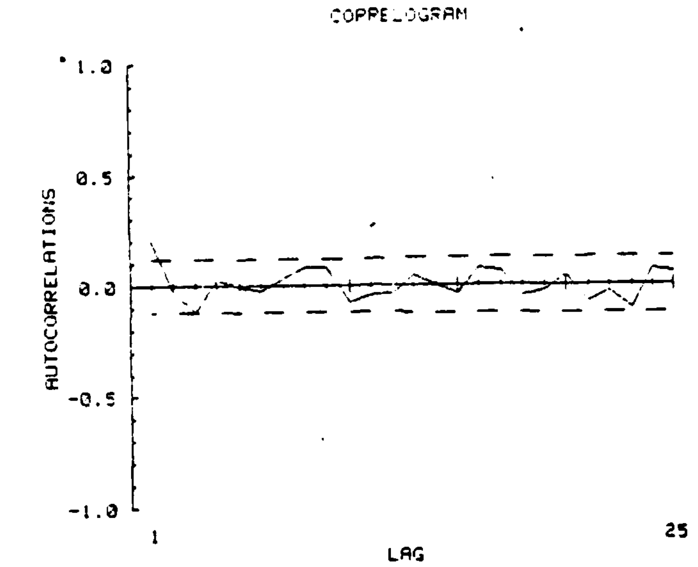


Figure 60. Correlogram of winter months only, logged rainfall anomalies from data set FL

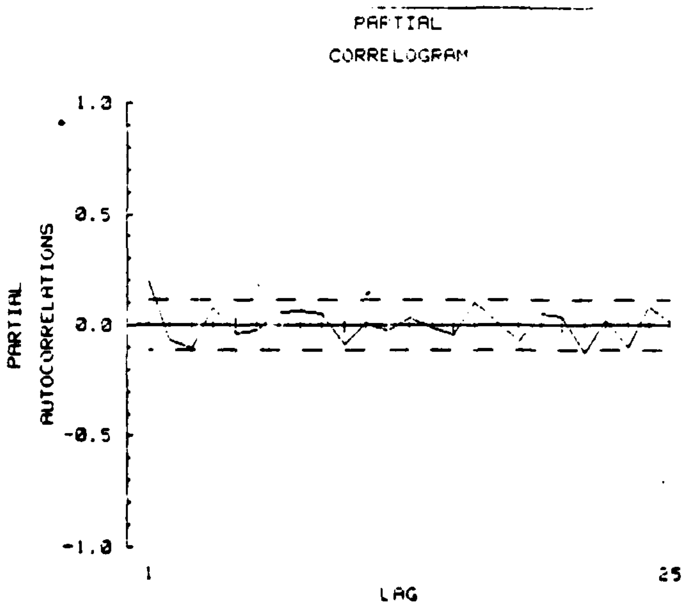
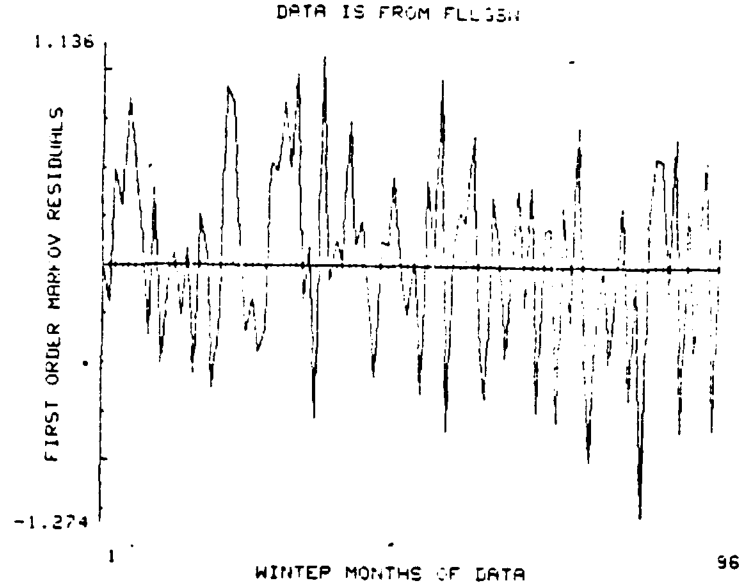


Figure 61. Partial correlogram of winter months only, logged rainfall anomalies from data set FL

These displays indicate a model like

$$\tilde{z}''_t = .199\tilde{z}''_{t-1} + a_t. \qquad \text{III.7}$$



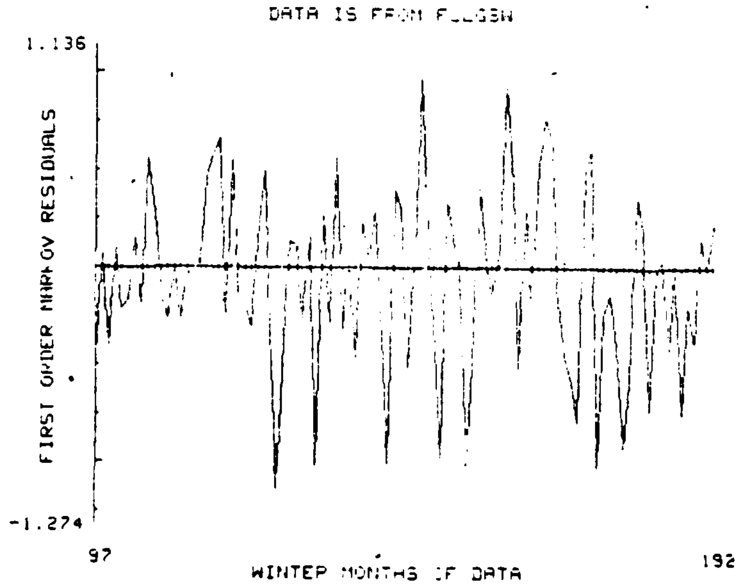


Figure 62a. Years 1 - 24, first order Markov residuals of logged rainfall anomalies for winter months only, data set FL

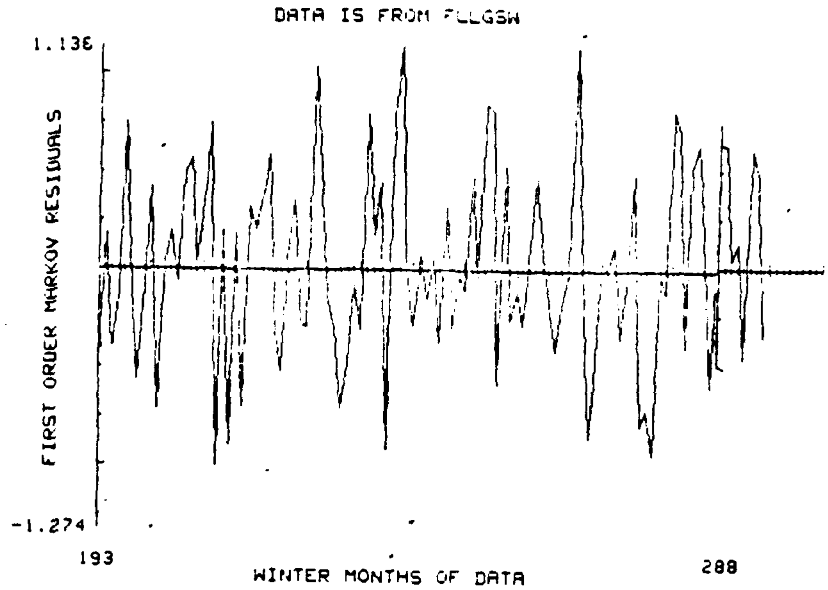


Figure 62b. Years 25 - 37, first order Markov residuals of logged rainfall anomalies for winter months only, data set FL

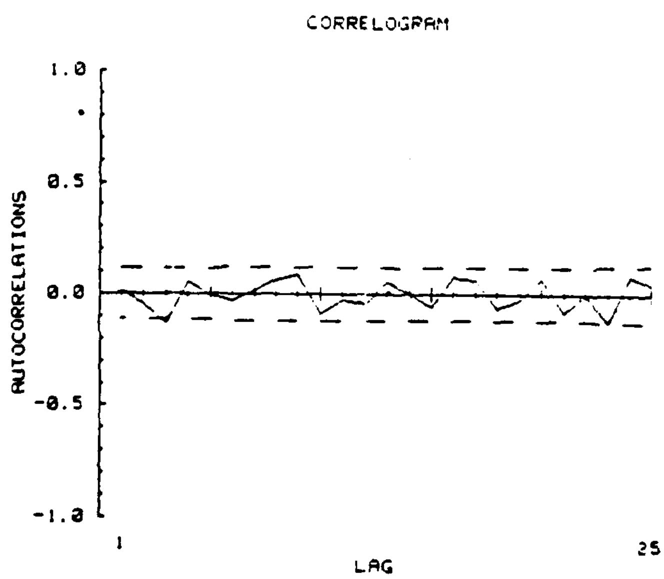


Figure 63. Correlogram of first order Markov residuals of lagged rainfall anomalies from winter months only, data set FL

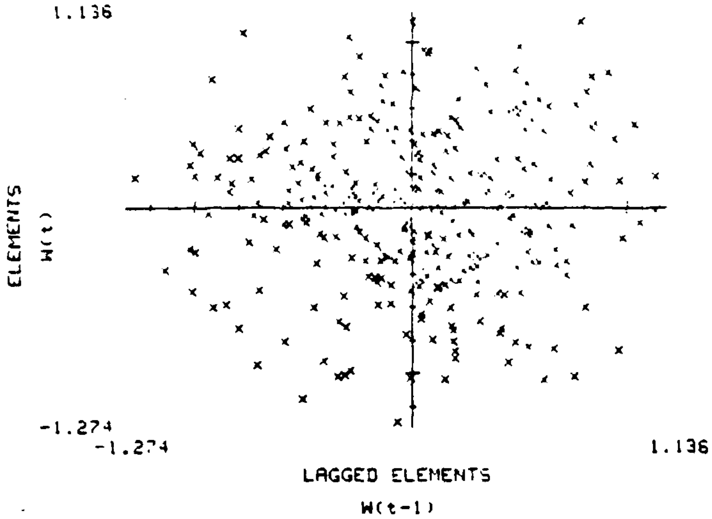


Figure 64. Lag one plot of first order Markov residuals from logged rainfall anomalies of winter months only, data set FL

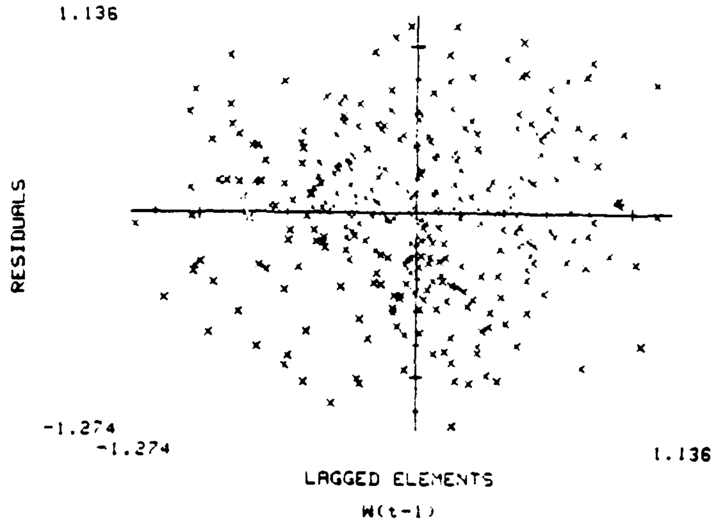


Figure 65. First order Markov residuals versus lag one data point from logged rainfall anomalies of winter months only, data set FL

TABLE 24

ACTUAL AND EXPECTED NUMBER OF TURNING POINTS AND ACTUAL AND EXPECTED PHASE FREQUENCIES FROM THE FIRST ORDER MARKOV RESIDUALS OF THE LOGGED RAINFALL ANOMALIES OF DATA SET FL

NUMBER OF TURNING POINTS = 209
E[P] = 1[illegible]6   V[P] = 15.902

PHASE LENGTHS

D	OBS.	E[*]
1	138	122.1
2	59	53.5
3	8	15.4
4	2	3.3
5	1	.6
6	0	.1
7	0	0.0
8	0	0.0

TABLE 25

GENERAL STATISTICS OF FIRST ORDER MARKOV RESIDUALS FROM LOGGED RAINFALL ANOMALIES OF WINTER MONTHS ONLY, DATA SET FL

Moments

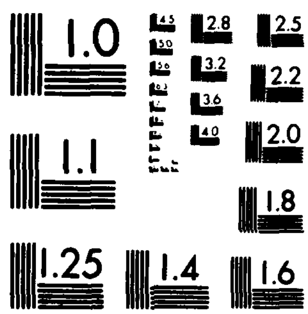
Mean	.000
Variance	.234
Skewness	-.079
Kurtosis	-.376

Percentiles

Minimum	-1.274
Lower Sixteenth	- .798
Lower Eight	- .570
Lower Quartile	- .315
Median	.011
Upper Quartile	.323
Upper Eight	.551
Upper Sixteenth	.748
Maximum	1.136

AD-A110 816 NAVAL POSTGRADUATE SCHOOL MONTEREY CA F/G 4/2
A STATISTICAL ANALYSIS OF MONTHLY RAINFALL FOR MONTEREY PENINSU--ETC(U)
MAR 81 D F DAVIS
UNCLASSIFIED NL

2 of 3



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

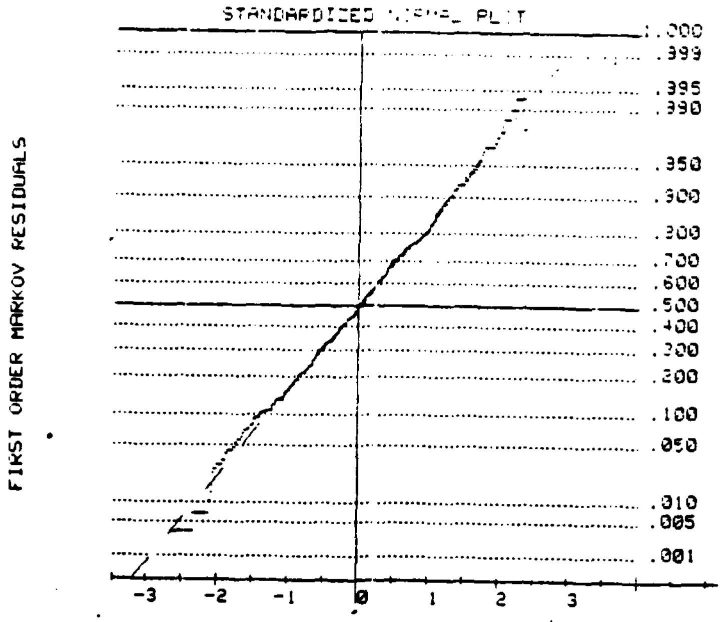


Figure 66. Standardized normal plot of first order Markov residuals from logged rainfall anomalies of winter months only, data set FL

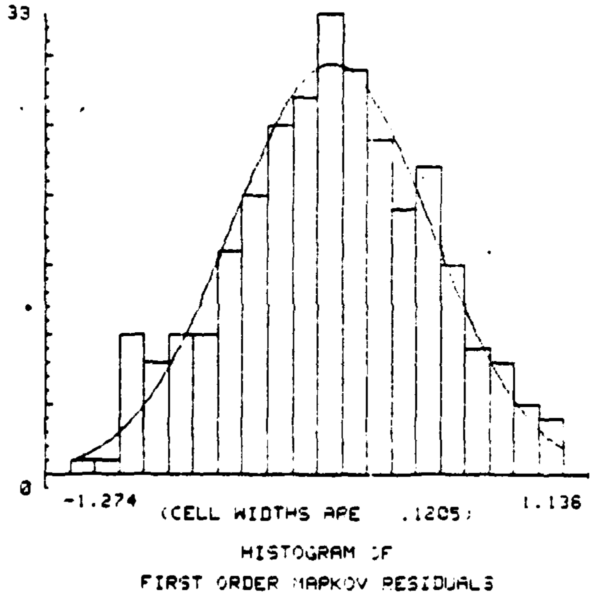


Figure 67. Histogram of first order Markov residuals from logged rainfall anomalies of winter months only, data set FL

The chi-squared was calculated at 15.35 for 17 degrees of freedom. This is a significance level of .570, thus indicating possible normality.

D. DATA SET SC

1. Twelve Month Series

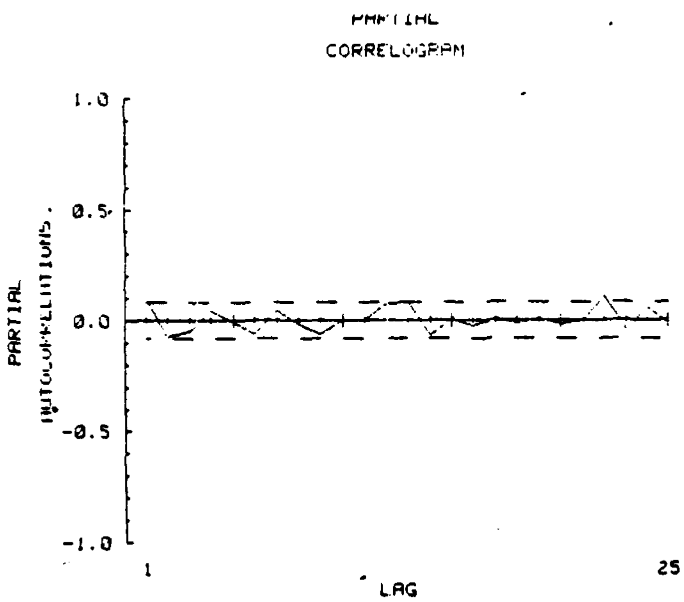


Figure 68. Partial correlogram of the logged anomalies for data set SC

TABLE 26

ESTIMATED PARTIAL AUTOCORRELATIONS
FOR LOGGED RAINFALL ANOMALIES OF
DATA SET SC

LAG	VALUE	LAG	VALUE
1	.096	14	-.066
2	-.075	15	.004
3	-.053	16	-.022
4	.046	17	.007
5	-.004	18	-.014
6	-.061	19	.004
7	.042	20	-.019
8	-.016	21	.002
9	-.061	22	.106
10	-.000	23	-.046
11	-.001	24	.059
12	.079	25	-.015
13	.084		

This information yields the model as

$$\tilde{Z}'_t = .096\tilde{Z}'_{t-1} + a_t. \qquad \text{III.12}$$

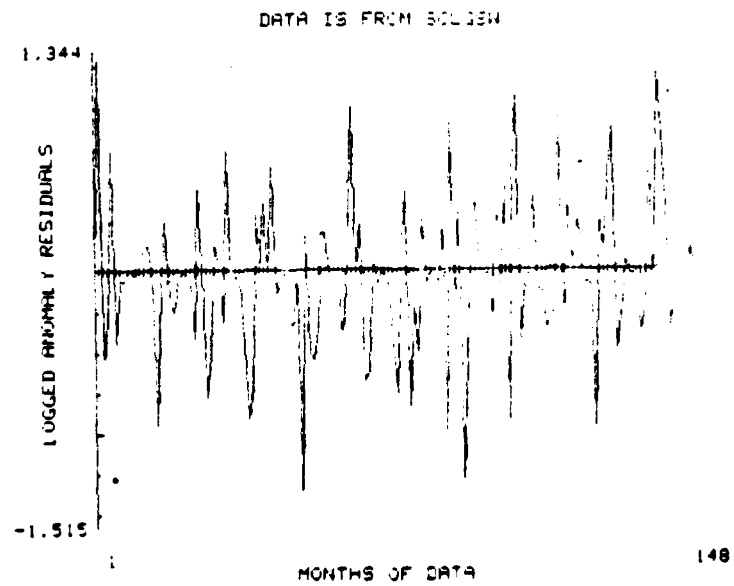


Figure 69a. First order Markov residuals from logged rainfall anomalies of data set SC. Months 1 - 148.

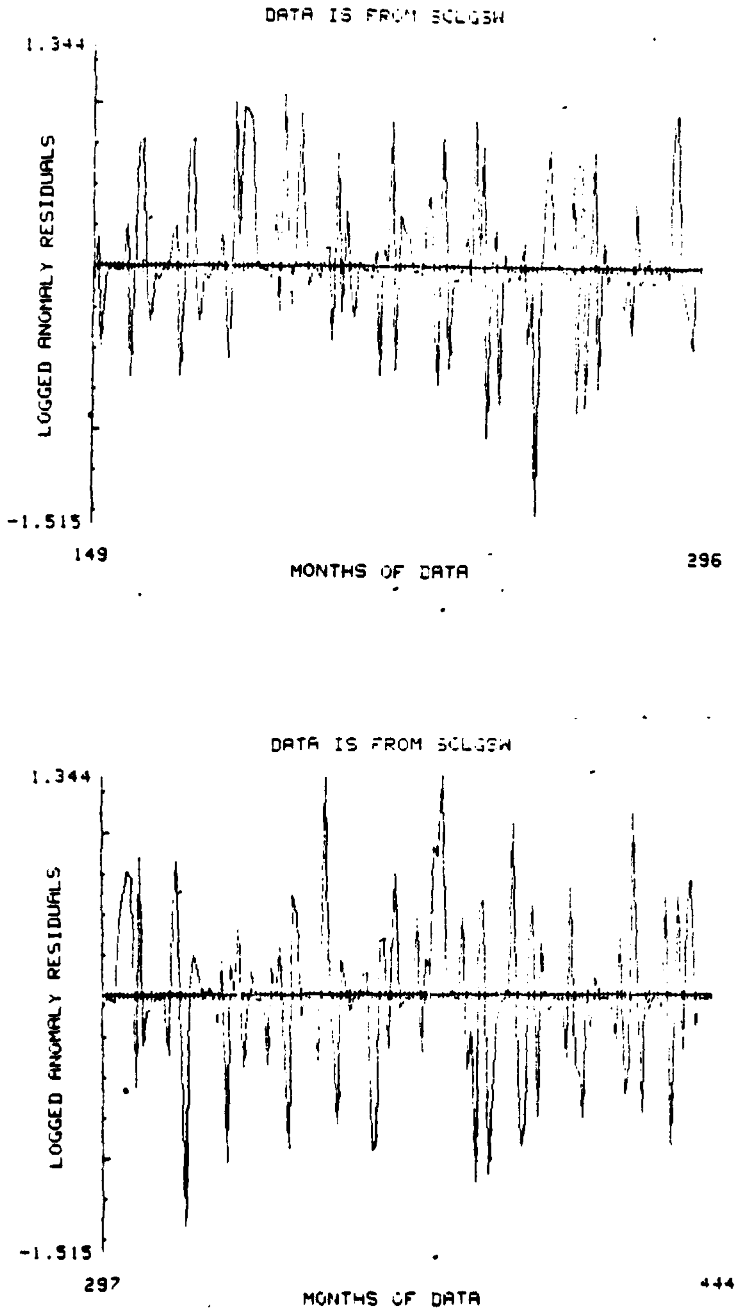


Figure 69b. First order Markov residual from logged rainfall anomalies of data set SC. Months 149 - 444

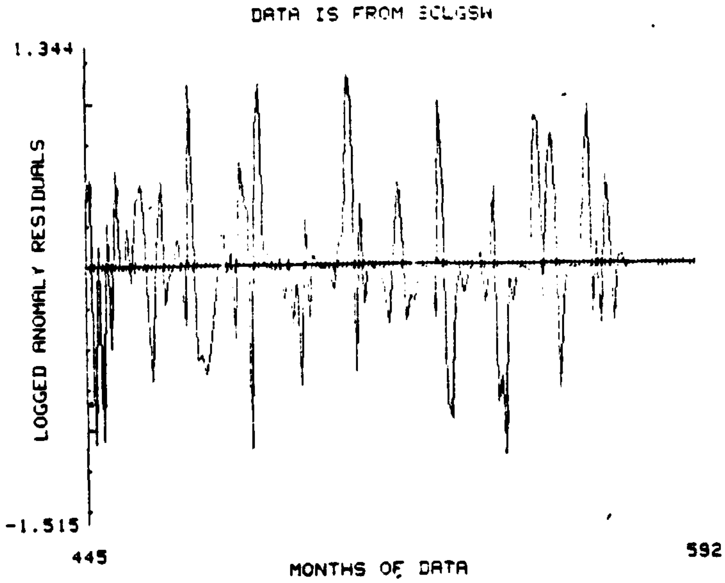


Figure 69c. First order Markov residual from logged rainfall anomalies of data set SC. Months 445 - 596.

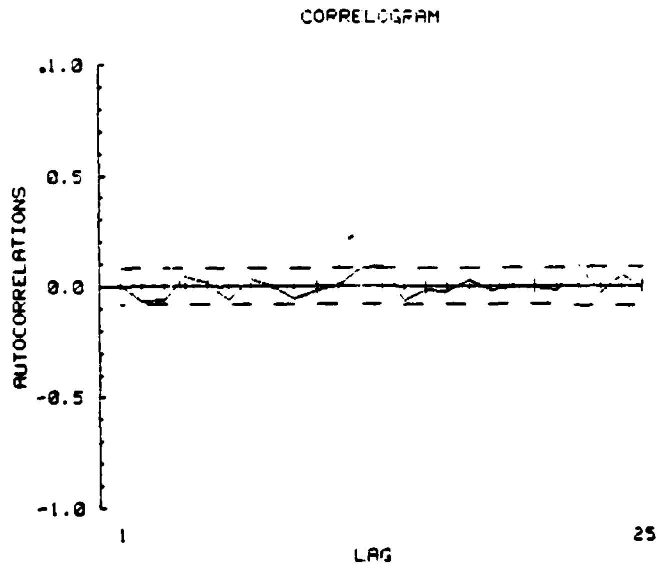


Figure 70. Autocorrelations of residuals from first order Markov process applied to the logged rainfall anomalies of data set SC

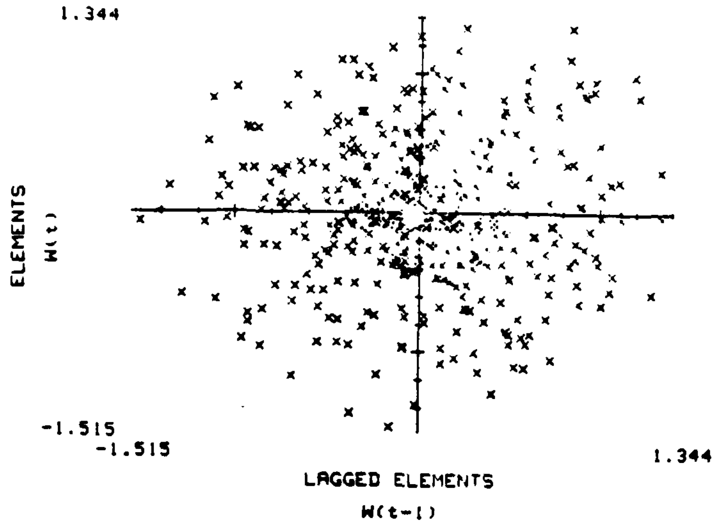


Figure 71. Lag one plot of first order Markov residuals from logged rainfall anomalies of data set SC

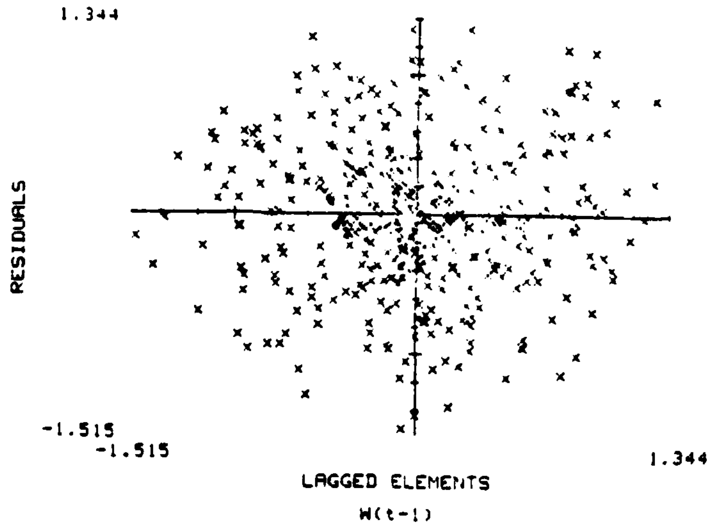


Figure 72. First order Markov residuals versus lag one data points from logged rainfall anomalies of data set SC

TABLE 27

ACTUAL AND EXPECTED NUMBER OF TURNING POINTS AND ACTUAL AND EXPECTED PHASE FREQUENCIES FROM THE FIRST ORDER MARKOV RESIDUALS OF DATA SET SC

NUMBER OF TURNING POINTS = 367
E[P] = 382.667   V[P] = 15.9

PHASE LENGTHS

D	OBS.	E[*]
1	226	238.8
2	95	104.9
3	34	30.1
4	8	6.6
5	2	1.2
6	1	.2
7	1	0.0
8	0	0.0
9	0	0.0
10	0	0.0
TOTALS	367	381.7

TABLE 28

GENERAL STATISTICS OF FIRST ORDER MARKOV RESIDUALS FROM LOGGED RAINFALL ANOMALIES OF DATA SET SC

Moments	
Mean	.000
Variance	.209
Skewness	.057
Kurtosis	.805
Percentiles	
Minimum	-1.515
Lower Sixteenth	- .744
Lower Eight	- .462
Lower Quartile	- .205
Median	- .030
Upper Quartile	.207
Upper Eight	.506
Upper Sixteenth	.783
Maximum	1.344

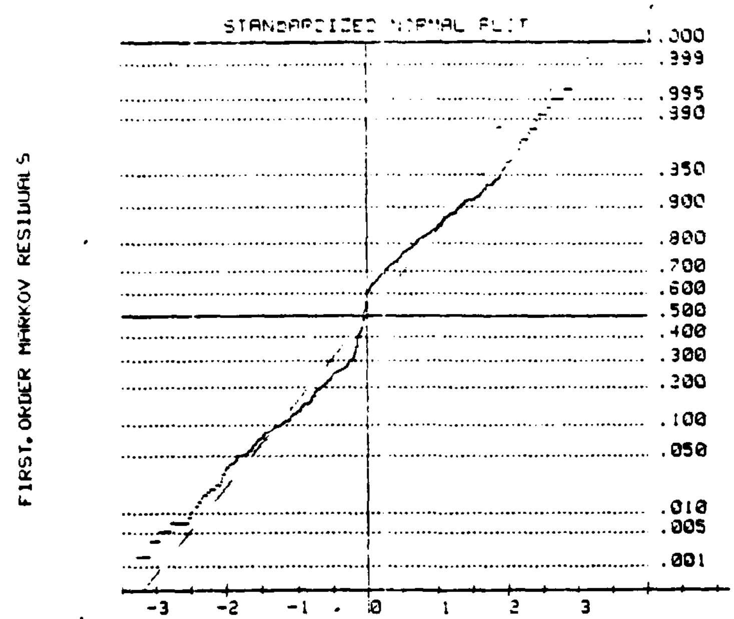


Figure 73. Standardized normal plot of first order Markov residuals from logged rainfall anomalies of data set SC

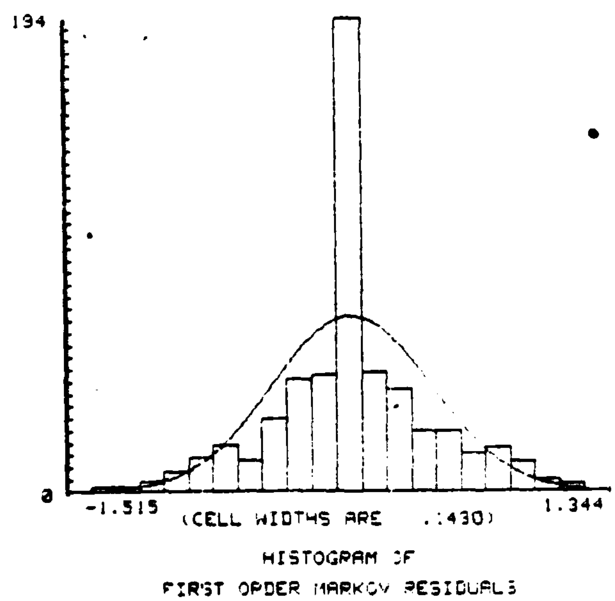


Figure 74. Histogram of first order Markov residuals from logged rainfall anomalies of data set SC

This data set yielded a chi-square value of 273.95 for 17 degrees of freedom. This is equivalent to a significance of zero plus.

2. <u>Winter Series</u>

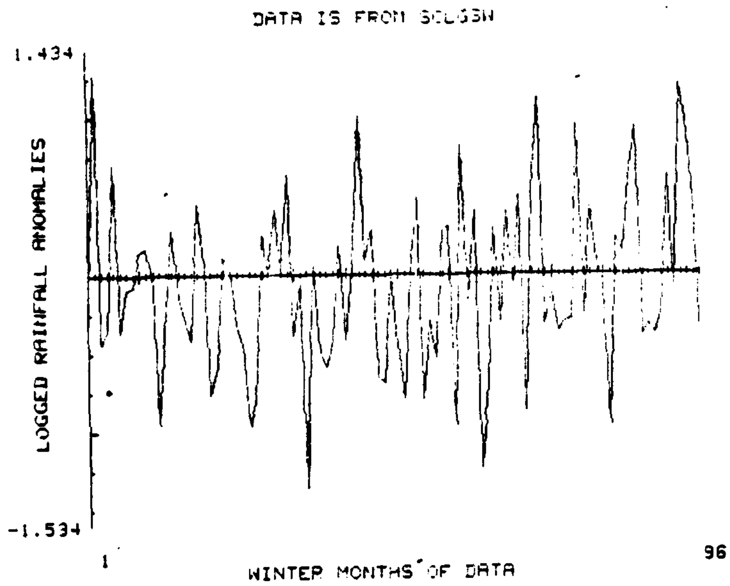
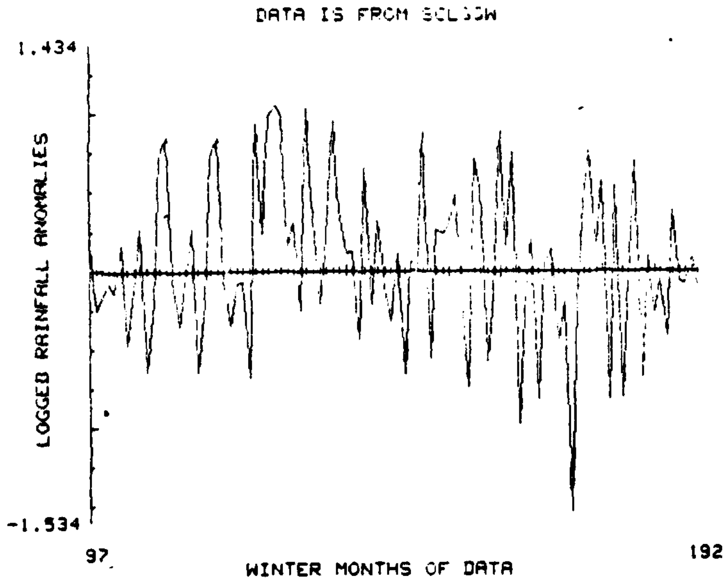


Figure 75a. Years 1 - 12 of winter months only, logged rainfall anomalies of data set SC.



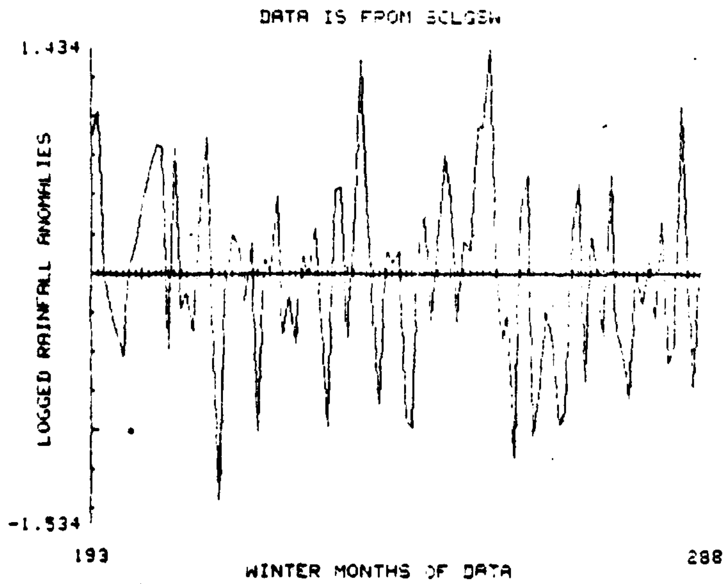


Figure 75b. Years 13 - 36 of winter months only, logged rainfall anomalies of data set SC

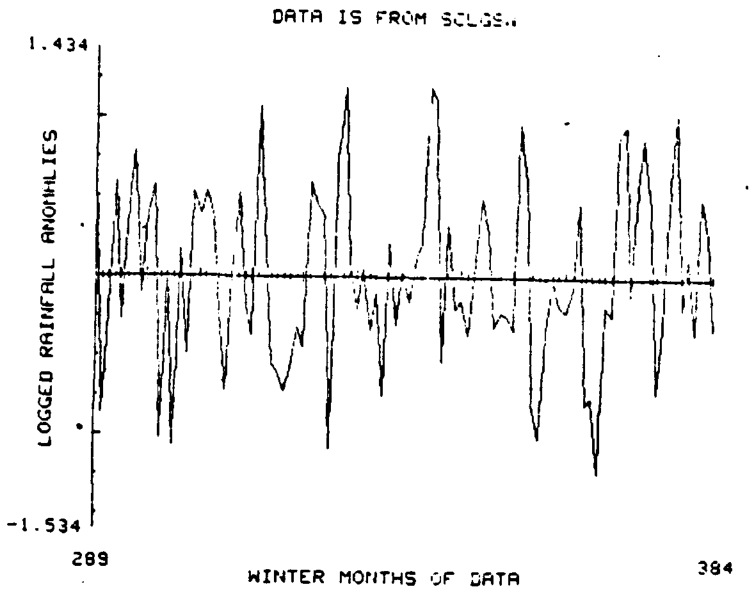


Figure 75c. Years 37 - 48 of winter months only, logged rainfall anomalies of data set SC

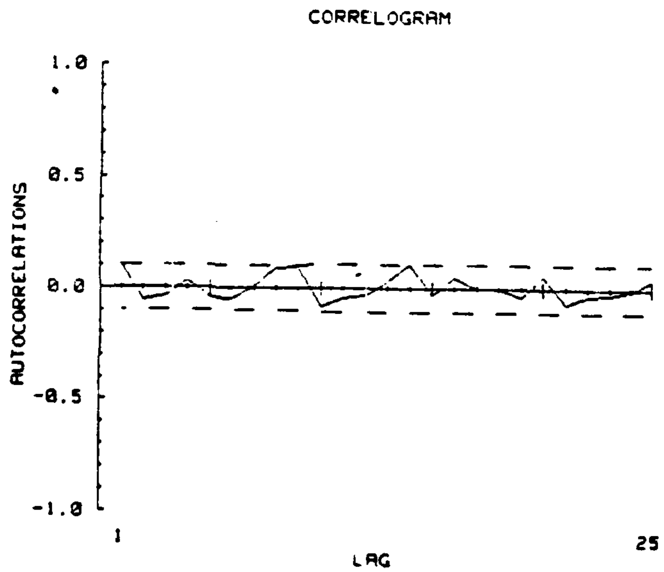


Figure 76. Correlogram of winter months only, logged rainfall anomalies from data set SC

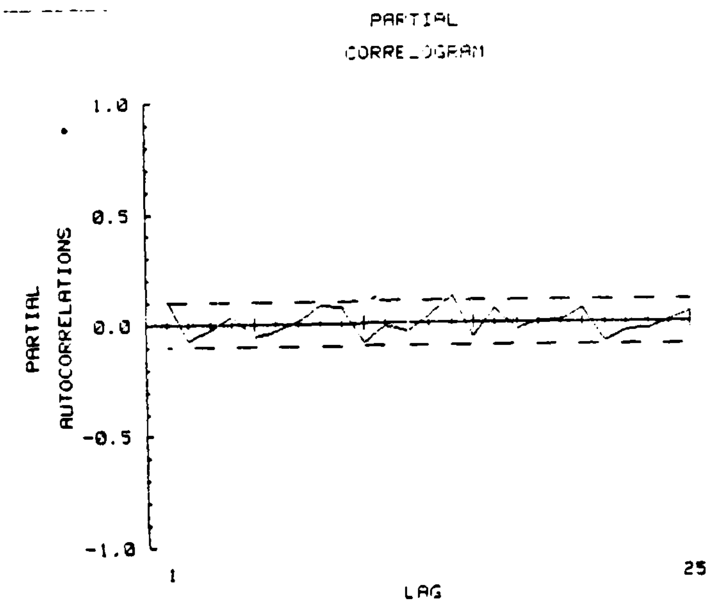
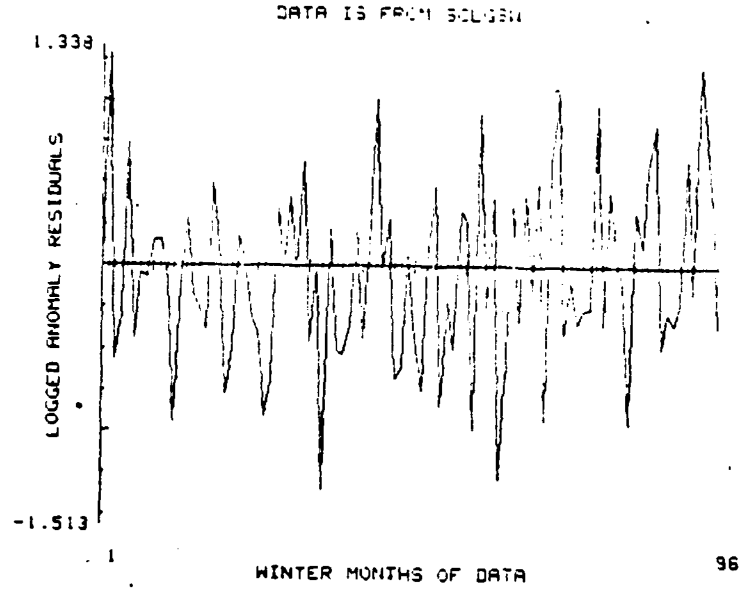


Figure 77. Partial correlogram of winter months only, logged rainfall anomalies from data set SC

This information indicates the model

$$\tilde{Z}_t = .107\tilde{Z}''_t + a_t. \qquad \text{III.13}$$



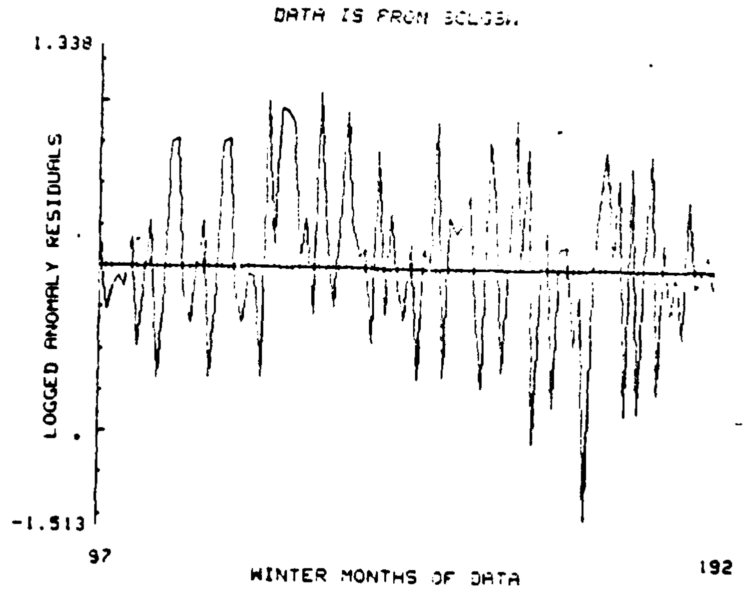
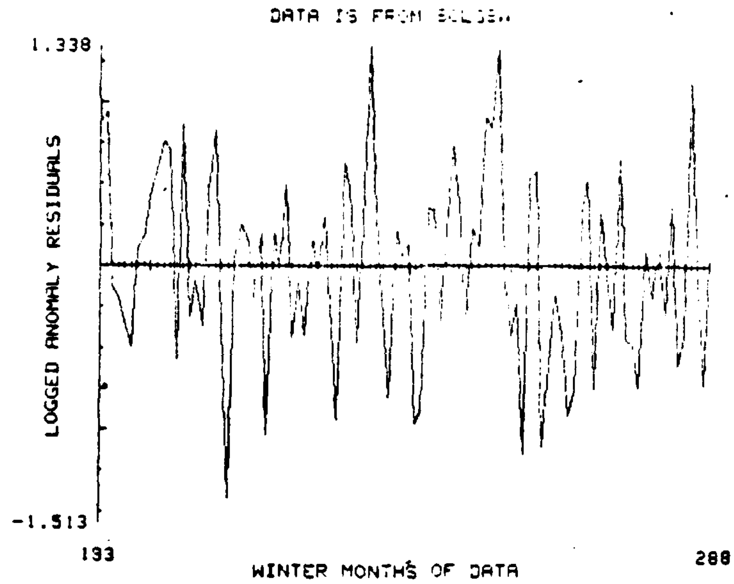


Figure 78a. Years - 24 first order Markov residuals of logged rainfall anomalies, for winter months only, data set SC



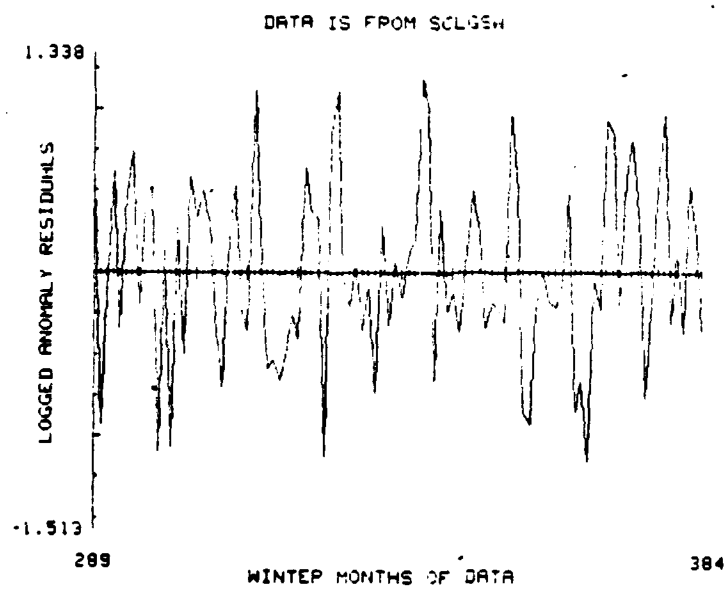


Figure 78b. Years 25 - 48, first order Markov residuals of logged rainfall anomalies for winter months only, data set SC

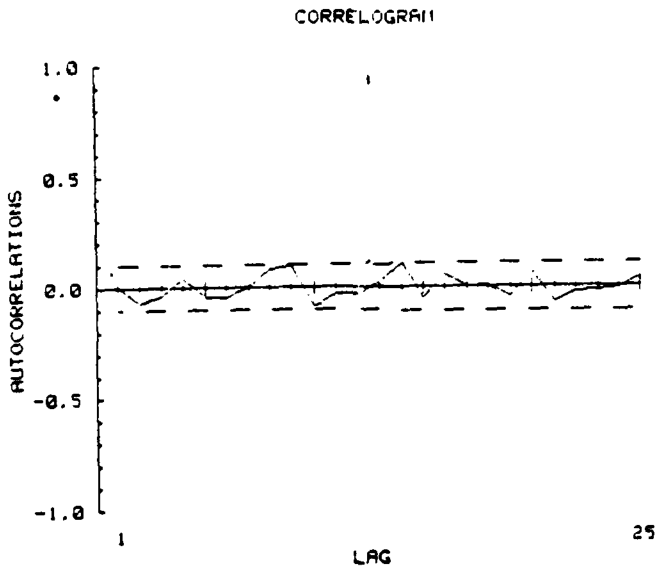


Figure 79. Correlogram of first order Markov residuals of logged rainfall anomalies from winter months only, data set SC.

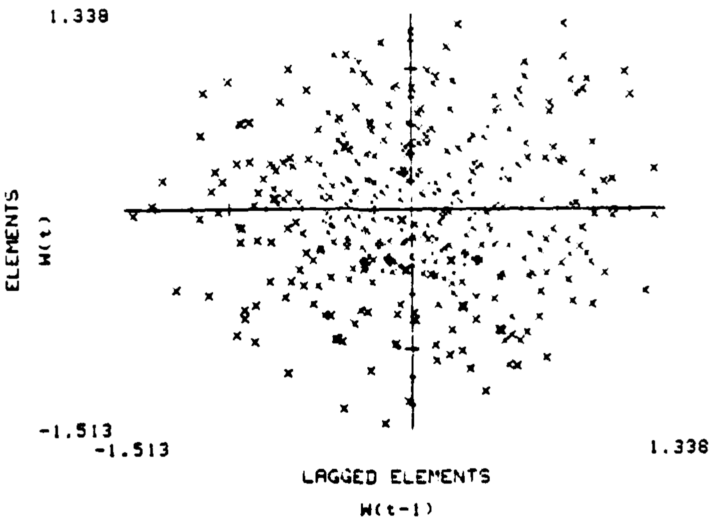


Figure 80. Lag one plot of first order Markov residuals from logged rainfall anomalies of winter months only, data set SC

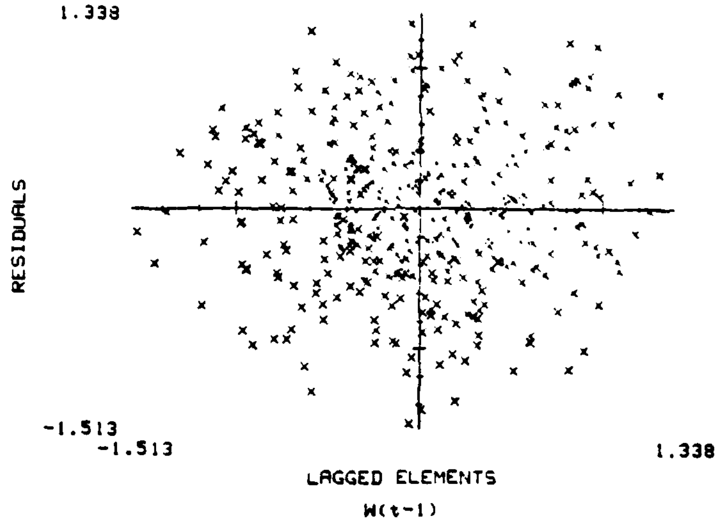


Figure 81. First order Markov residuals versus lag one data point from logged rainfall anomalies of winter months only, data set SC

TABLE 29

ACTUAL AND EXPECTED NUMBER OF TURNING POINTS AND ACTUAL AND EXPECTED PHASE FREQUENCIES FROM THE FIRST ORDER MARKOV RESIDUALS OF THE LOGGED RAINFALL ANOMALIES OF THE WINTER MONTHS ONLY, DATA SET SC

NUMBER OF TURNING POINTS = 32
E[P] = 30.667 V[P] = 15.39

PHASE LENGTHS

D	OBS.	E[*]
1	21	18.8
2	8	8.1
3	3	2.3
4	0	.5
5	0	0.0
6	0	0.0
7	0	0.0
8	0	0.0
9	0	0.0
10	0	0.0
TOTALS	32	29.7

TABLE 30

GENERAL STATISTICS OF FIRST ORDER MARKOV RESIDUALS FROM LOGGED RAINFALL ANOMALIES OF WINTER MONTHS ONLY, DATA SET SC

Moments	
Mean	.000
Variance	.305
Skewness	.015
Kurtosis	-.363

Percentiles	
Minimum	-1.513
Lower Sixteenth	- .872
Lower Eight	- .663
Lower Quartile	- .359
Median	- .026
Upper Quartile	.355
Upper Eight	.682
Upper Sixteenth	.882
Maximum	1.338

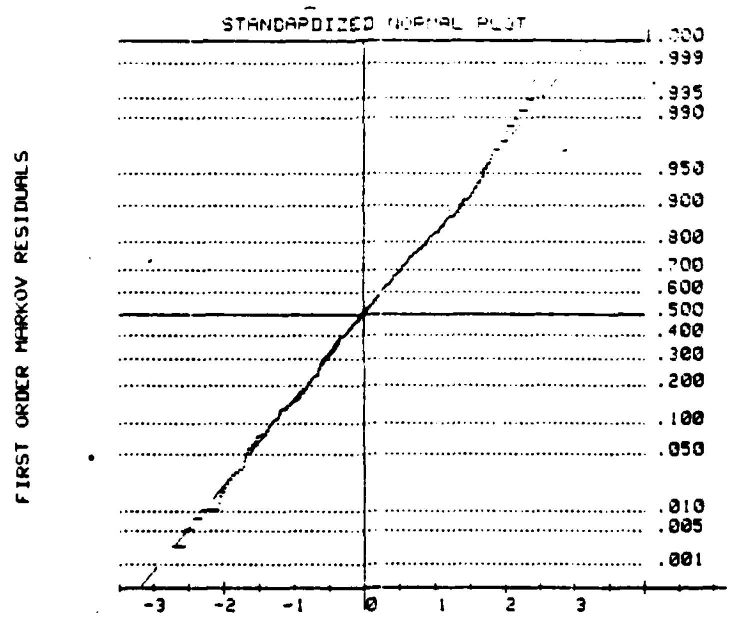


Figure 82. Standardized normal plot of first order Markov residuals from logged rainfall anomalies of winter months only, data set SC

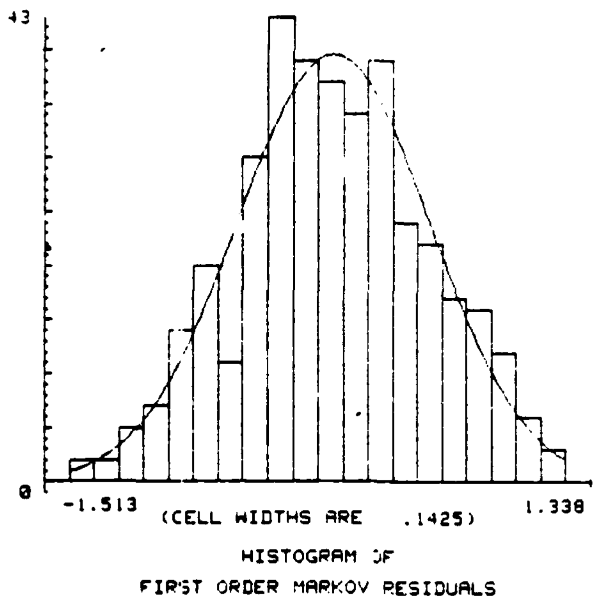


Figure 83. Histogram of first order Markov residuals from logged rainfall anomalies of winter months only, data set SC

The chi-square was calculated at 16.60 for 17 degrees of freedom. This is significant at the .482 level thus indicating probable normality.

IV. VALIDATION OF FIRST ORDER MARKOV MODELS

A. THEORY

The general model proposed by a first order Markov process is, as stated before;

$$\tilde{Z}_t = \rho\tilde{Z}_t + a_t \qquad \text{IV.1}$$

where $\{a_t\}$ are distributed iid $N(0, \sigma_a^2)$. To validate this model, preferably independent data should be subjected to the model, and an analysis of the residual, or forecast errors, made.

As stated previously, years 1975 through 1980 were reserved for the purpose of validation. The method of validation was to use the model to construct a series of one step ahead forecasts. Let $e_t(1)$ be the error in a forecast of time t+1 from the model at time t. Then the minimum mean squared error forecast (see Box and Jenkins) is;

$$e_t(1) = \tilde{Z}_t - \hat{\rho}\tilde{Z}_{t-1} \qquad \text{IV.2}$$

If the model is correct, the sequence $\{e_t(1)\}$ will be independent normally distributed with mean zero and variance σ_a^2. In the following sections the models are applied to the reserved data sets (which may also be found in the appendixes), and these forecast errors are calculated. The forecast errors are then analyzed to determine if

(1) The errors are serially independent

(2) The errors are distributed as normal random variables with mean zero, and variance σ_a^2.

Since the residual analysis of the twelve month model already indicates a poor fit, the twelve month model will not be validated. Only the winter month models will be checked for validity.

B. DATA SET RN

Figures 84 (Raw data) and 85 (Logged anomalies) display the reserved data set. The logged anomalies were formed by removing the means of the analyzed data, Table 4, not the means of the logged reserved data. This was done to remove any bias from the validation.

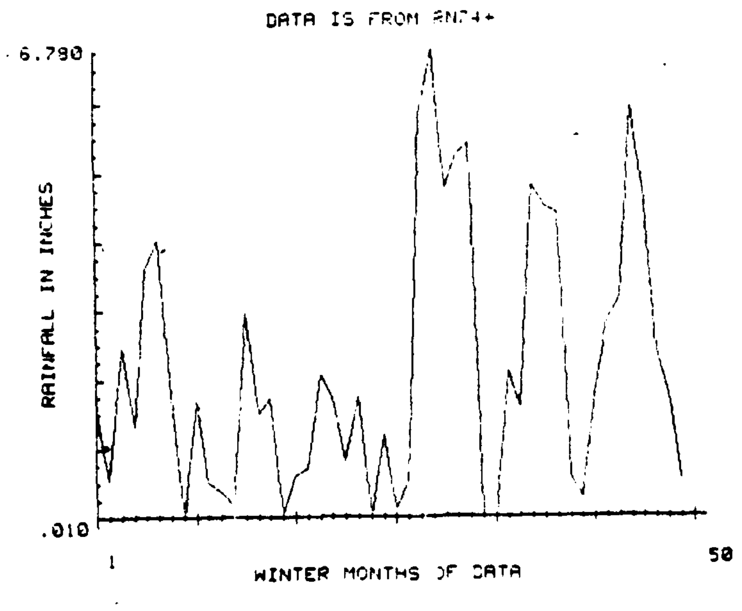


Figure 84. Reserved rainfall data for data set RN

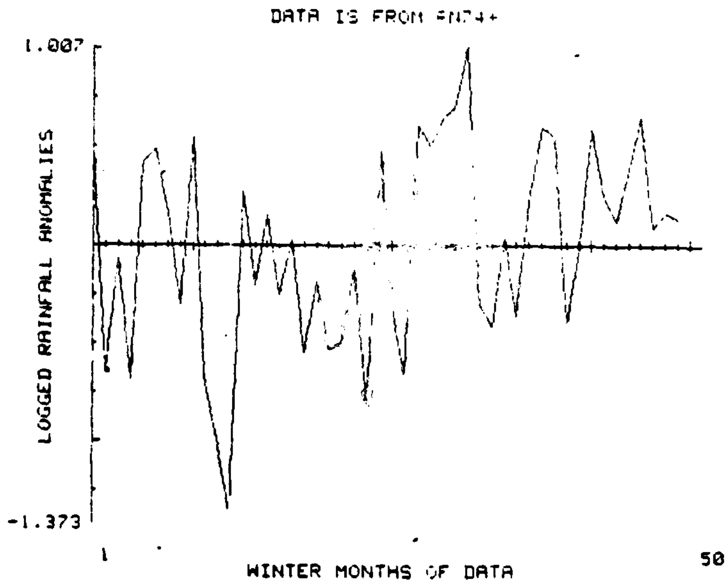


Figure 85. Logged rainfall anomalies of reserved data set RN

The forecast errors, Figure 86, their correlogram, Figure 87, and independence tests, Table 31 indicate that the errors are indeed independent.

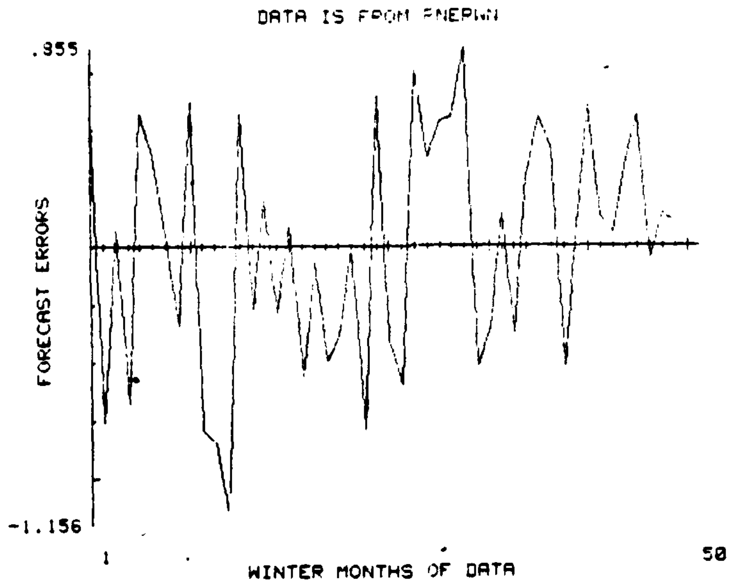


Figure 86. Forecast errors from first order Markov model applied to winter months of reserved data set RN

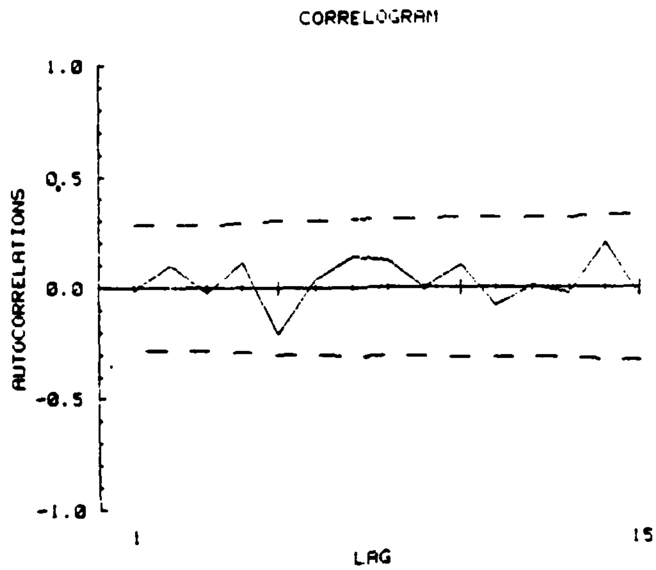


Figure 87. Correlogram of forecast errors from first order Markov model applied to winter months of reserved data set RN

TABLE 31

ACTUAL AND EXPECTED NUMBER OF TURNING POINTS AND ACTUAL AND EXPECTED PHASE FREQUENCEIS FOR THE FORECAST ERRORS OF THE FIRST ORDER MARKOV MODEL APPLIED TO THE WINTER MONTHS OF RESERVED DATA SET RN

NUMBER OF TURNING POINTS = 32
E[P] = 30.667 V[P] = 15.39

PHASE LENGTHS

D	OBS.	E[*]
1	21	18.8
2	8	8.1
3	3	2.3
4	0	.5
5	0	0.0
6	0	0.0
7	0	0.0
8	0	0.0
9	0	0.0
10	0	0.0
TOTALS	32	29.7

The normality of the forecast errors is addressed by Table 32 (Statistics), Figure 88 (Normal plot), 89 (Histogram), and a simple chi-squared test. The chi-squared was calculated as 7.82 with 5 degrees of freedom which is significant at the .167 level. However, the normality of the errors is somewhat questionable due to the other displays.

TABLE 32

GENERAL STATISTICS OF FORECAST ERRORS FROM THE FIRST ORDER MARKOV MODEL APPLIED TO THE WINTER MONTHS OF RESERVED DATA SET RN

Moments

Mean	.001
Variance	.259
Skewness	-.276
Kurtosis	-.955

Percentiles

Minimum	-1.156
Lower Sixteenth	- .791
Lower Eight	- .634
Lower Quartile	- .397
Median	.069
Upper Quartile	.409
Upper Eight	.569
Upper Sixteenth	.614
Maximum	.855

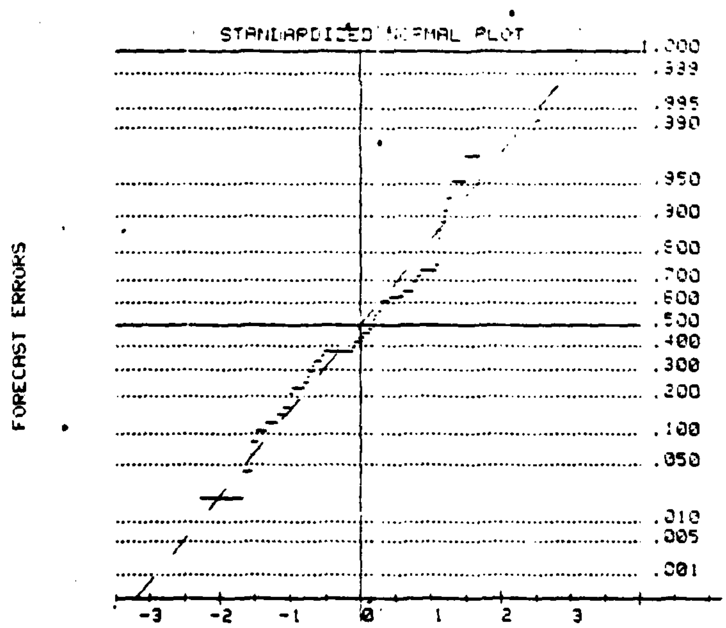


Figure 88. Standardized normal plot of forecast errors from the first order Markov model applied to the winter months of reserved data set RN

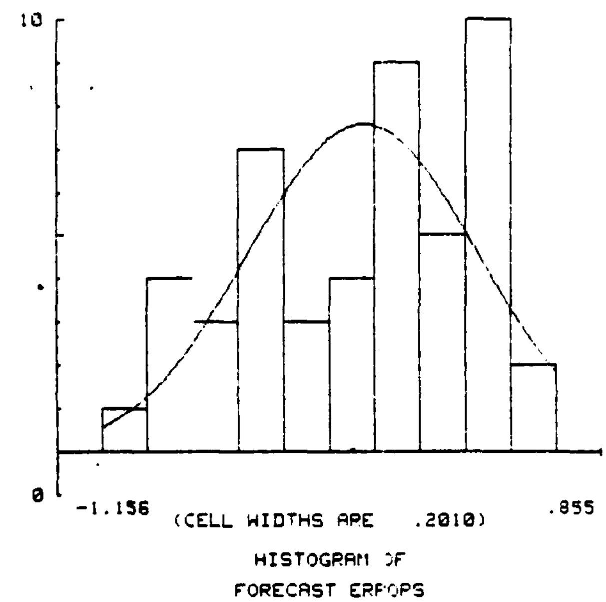


Figure 89. Histogram of forecast errors from the first order Markov model applied to the winter months of reserved data set RN

C. DATA SET FL

As before, the similarity of results for the different data set allows the analysis to be portrayed using the displays only.

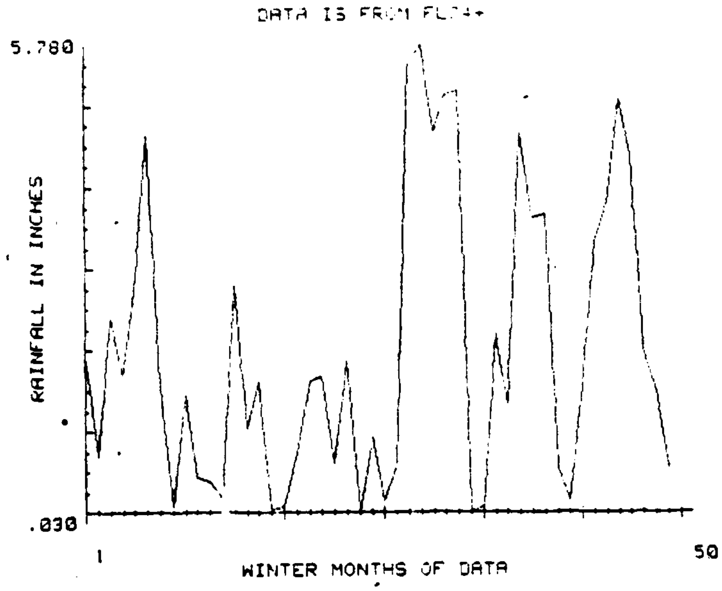


Figure 90. Reserved rainfall data for data set FL

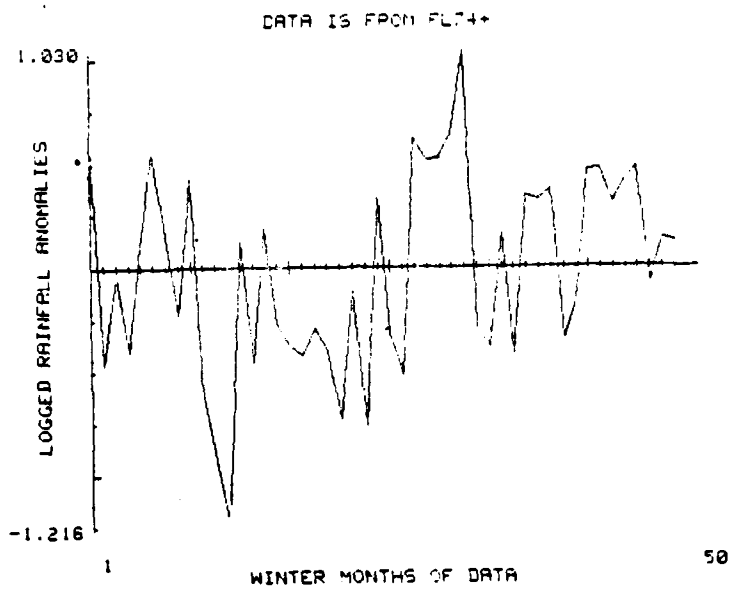


Figure 91. Logged rainfall anomalies of reserved data set FL.

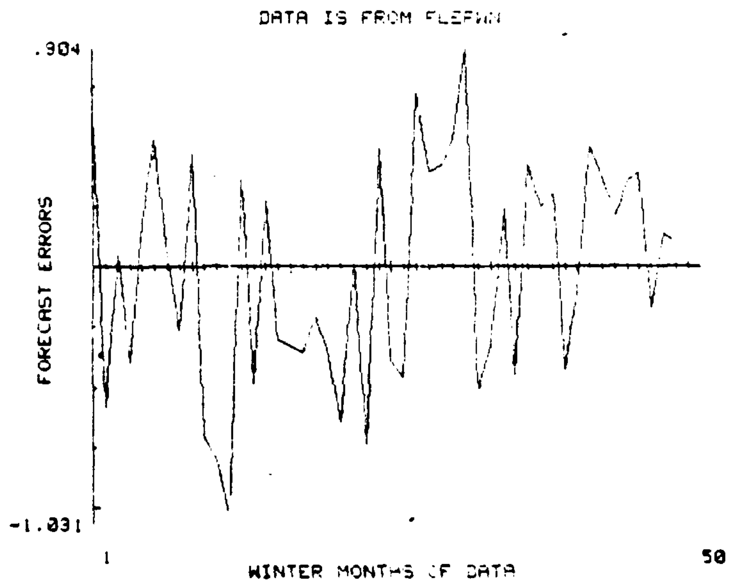


Figure 92. Forecast errors from the first order Markov model applied to the winter months of reserved data set FL

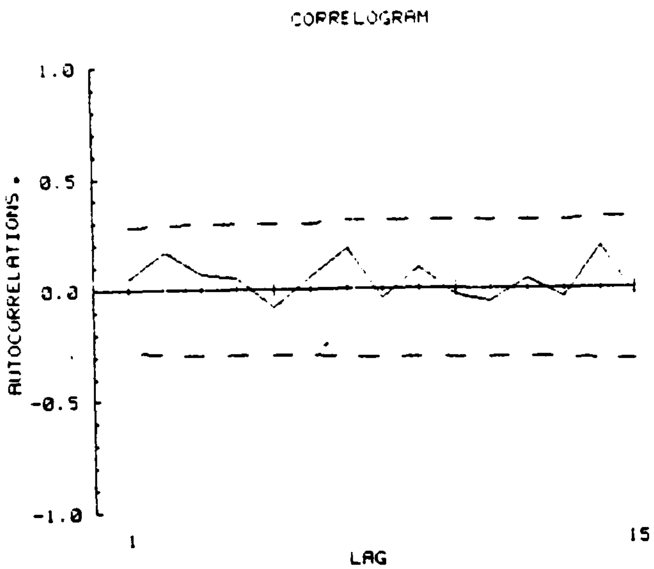


Figure 93. Correlogram of forecast errors from first order Markov model applied to the winter months of reserved data set FL

TABLE 33

ACTUAL AND EXPECTED NUMBER OF TURNING POINTS AND ACTUAL AND EXPECTED PHASE FREQUENCIES FROM THE FORECAST ERRORS OF THE FIRST ORDER MARKOV MODEL APPLIED TO THE WINTER MONTHS OF RESERVED DATA SET FL

NUMBER OF TURNING POINTS = 34

E[P] = 30.667 V[P] = 15.396

PHASE LENGTHS

D	OBS.	E[*]
1	24	18.8
2	8	8.1
3	2	2.3
4	0	.5
5	0	0.0
6	0	0.0
7	0	0.0
8	0	0.0
9	0	0.0
10	0	0.0
TOTALS	34	29.7

TABLE 34

GENERAL STATISTICS OF FORECAST ERRORS FROM THE FIRST ORDER MARKOV MODEL APPLIED TO THE WINTER MONTHS OF RESERVED DATA SET FL

Moments

Mean	-.015
Variance	.215
Skewness	-.155
Kurtosis	-.970

Percentiles

Minimum	-1.031
Lower Sixteenth	- .724
Lower Eight	- .549
Lower Quartile	- .396
Median	.049
Upper Quartile	.369
Upper Eight	.479
Upper Sixteenth	.531
Maximum	.904

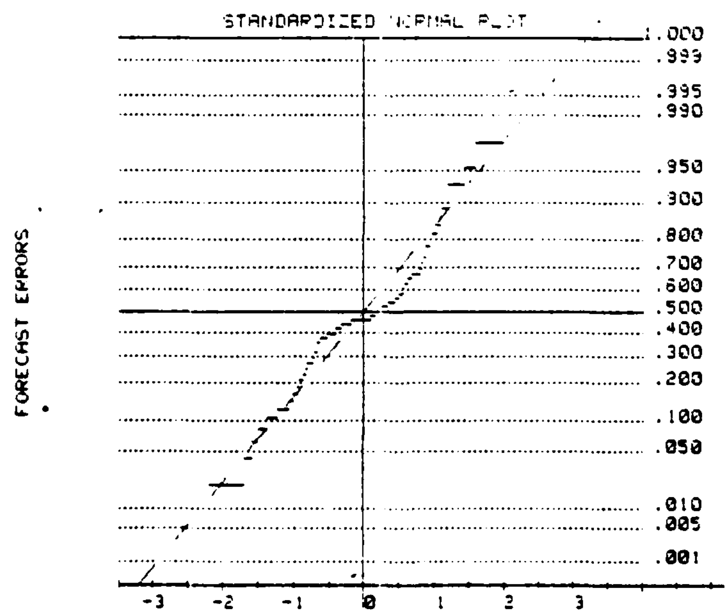


Figure 94. Standardized normal plot of forecast errors from the first order Markov model applied to the winter months of reserved data set FL

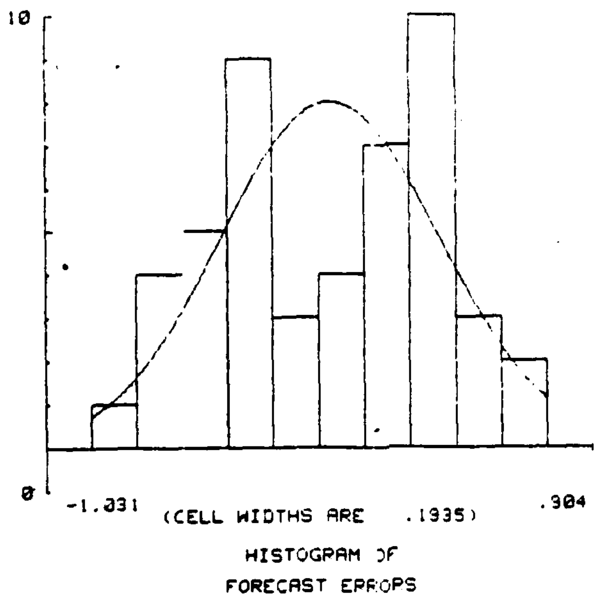


Figure 95. Histogram of froecast errors from the first order Markov model applied to the winter months of reserved data set FL

The chi-squared statistic was calculated as 12.58 with 7 degrees of freedom, thus yielding a significance level of 0.083. This statistic and the displays imply that the data are only marginally normal if at all.

D. DATA SET SC

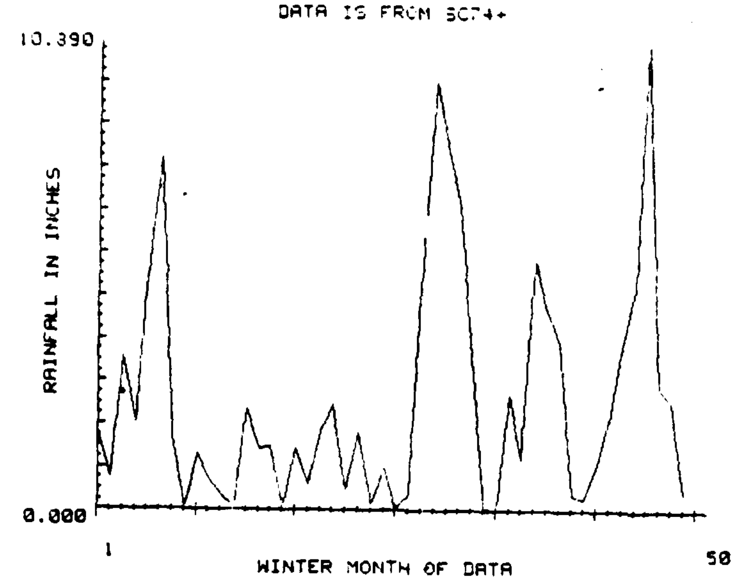


Figure 96. Reserved rainfall for data set SC

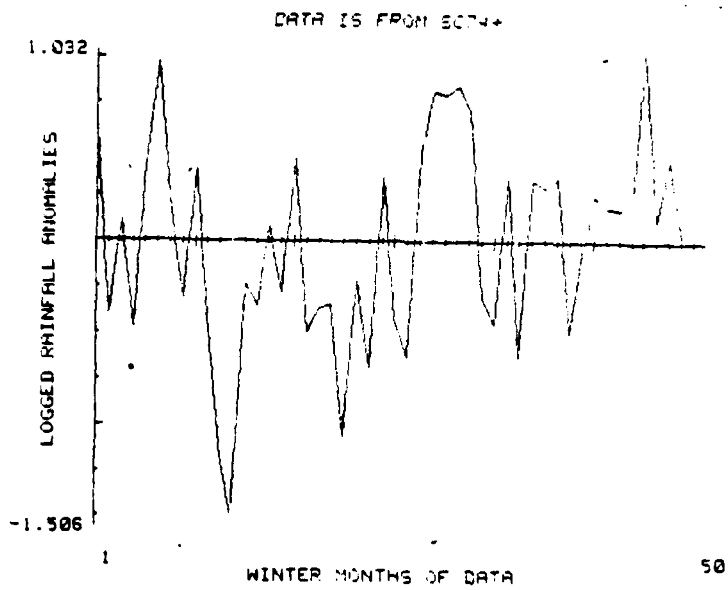


Figure 97. Logged anomalies of reserved data set SC

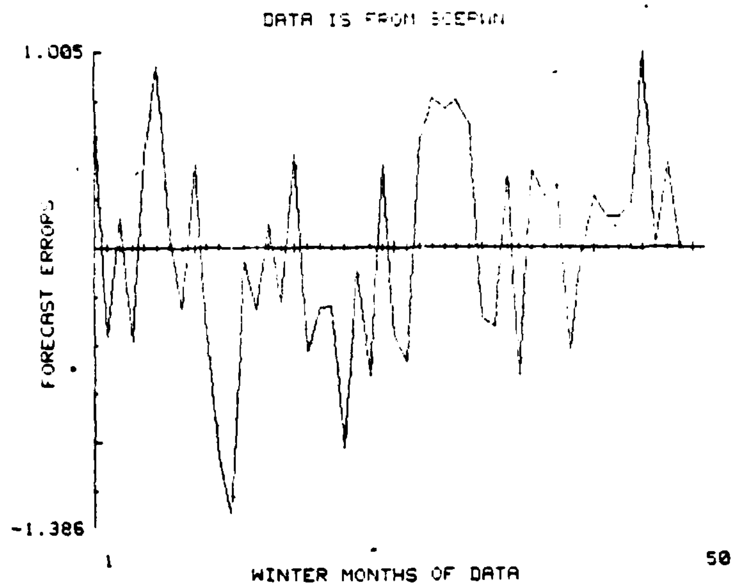


Figure 98. Forecast errors from first order Markov model applied to the winter months of reserved data set SC

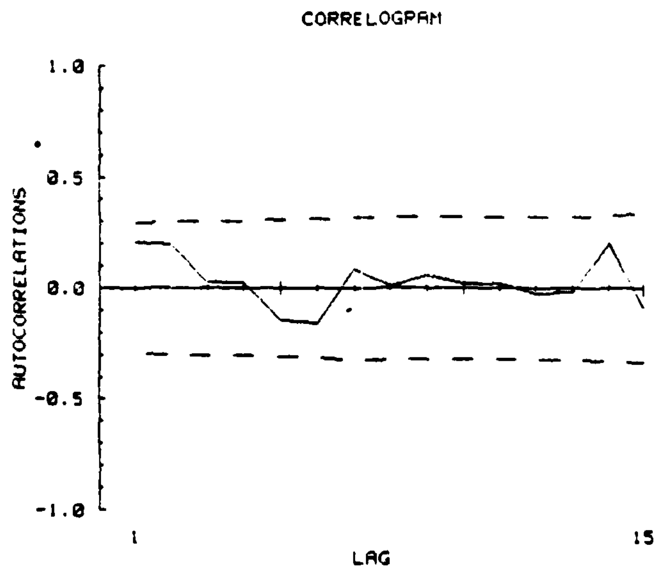


Figure 99. Correlogram of forecast errors from first order Markov model applied to the winter months of reserved data set SC

TABLE 35

ACTUAL AND EXPECTED NUMBER OF TURNING POINTS AND ACTUAL AND EXPECTED PHASE FREQUENCES FROM THE FORECAST ERRORS OF THE FIRST ORDER MARKOV MODEL APPLIED TO THE WINTER MONTHS OF RESERVED DATA SET SC

NUMBER OF TURNING POINTS = 34

E[P] = 30.667 V[P] = 15.396

PHASE LENGTHS

D	OBS.	E[*]
1	24	18.8
2	8	8.1
3	2	2.3
4	0	.5
5	0	0.0
6	0	0.0
7	0	0.0
TOTALS	34	29.7

TABLE 36

GENERAL STATISTICS OF FORECAST ERRORS FROM THE FIRST ORDER MARKOV MODEL APPLIED TO THE WINTER MONTHS OF RESERVED DATA SET SC

Moments

Mean	-.003
Variance	.296
Skewness	-.298
Kurtosis	-.413

Percentiles

Minimum	-1.386
Lower Sixteenth	- .846
Lower Eight	- .559
Lower Quartile	- .434
Median	.031
Upper Quartile	.414
Upper Eight	.588
Upper Sixteenth	.735
Maximum	1.005

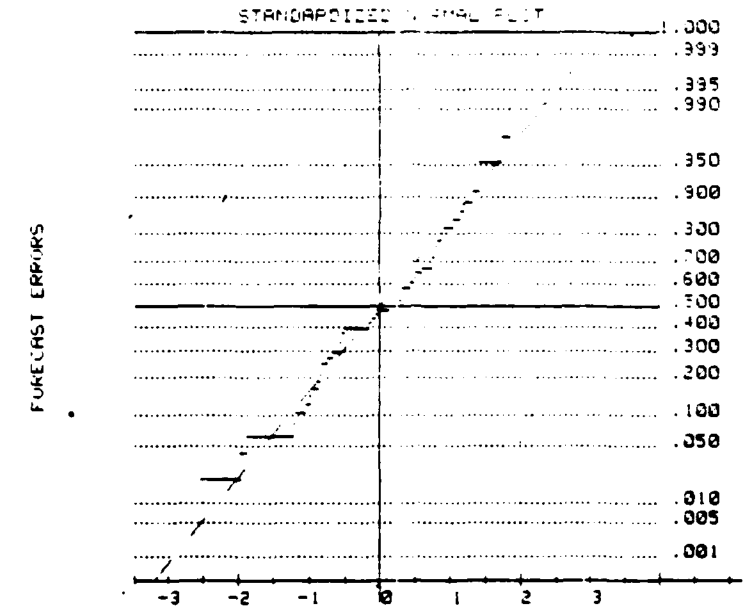


Figure 100. Standardized normal plot of forecast errors from the first order Markov model applied to the winter months of reserved data set SC

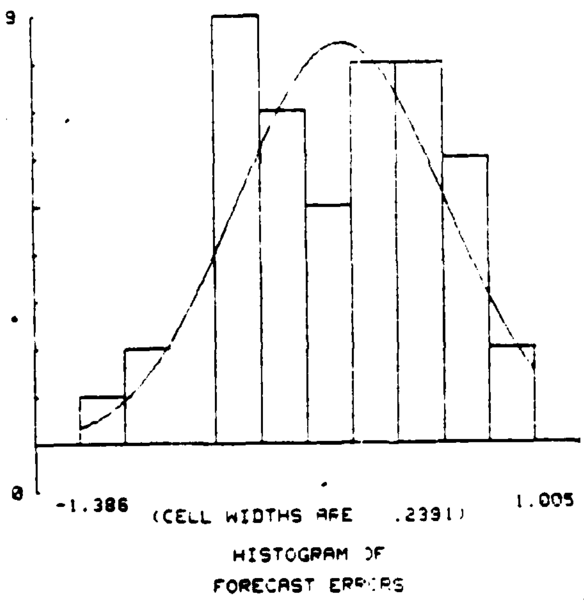


Figure 101. Histogram of forecast errors from the first order Markov model applied to the winter months of reserved data set SC

The chi-squared statistic was calculated as 9.71 with 7 degrees of freedom, thus yielding a significance level of .205.

E. CONCLUSIONS

The application of a Markovian model was indicated by the apparent dependence of adjacent months and the apparent lack of dependence at any other lag. The preceeding subsections, however, indicate that the first order Markovian model is weak at best.

The structure of the data, visually, still points toward some sort of underlying order. The following sections attempt to discover this order.

V. 2x2 TABLES

A. THEORY

As seen in sections III and IV, the classical ARMA time series approach does not seem to adequately describe the data. Another technique used to explore possible relationships is the 2x2 contingency tables.

The idea to be explored is whether or not some subset of the data, to be called the control, may be used to predict in some way the behavior of another subset of the data, to be called the complement. Here, the data are reduced from monthly observations to yearly observations as described below.

Let $\underline{X}$ be the subset of a year, to be called the control, and let $\underline{Y}$ be the subset to be called the complement. It is necessary that $\underline{X} \cap \underline{Y} = \emptyset$; that is, the intersection of these two sets is empty. The data are then compared for some quality in $\underline{X}$ and for some quality in $\underline{Y}$. The question is then: does the presence (or absence) of the quality in $\underline{X}$ affect the presence (or absence) of the quality in $\underline{Y}$? An example of a typical table is shown below in Figure 102.

		COMPLEMENT $\underline{Y}$ 1	2	
CONTROL $\underline{X}$	1	n_{11}	n_{12}	$n_{1\cdot}$
	2	n_{21}	n_{22}	$n_{2\cdot}$
		$n_{\cdot 1}$	$n_{\cdot 2}$	$n_{\cdot\cdot}$

Figure 102. Typical 2x2 contingency table

The table elements, n_{ij}, represents the number of years which display quality i in the control and quality j in the complement. The marginal entries $n_{i\cdot}$ and $n_{\cdot j}$ represents the numbers of years for which the control has quality i and the number of years the complement has quality j respectively. The overall number of years, $n_{\cdot\cdot}$, is in the lower right of the table.

Brownlee [Ref. 6] contains a very good discussion of the theory and use of 2x2 contingency tables. Using the notation of Brownlee, let θ_{ij} be the probability that any given year

will have a control quality i and a complement quality j. Then estimates of the θ_{ij}'s are

$$\hat{\theta}_{ij} = n_{ij}/n_{..}$$
$$\hat{\theta}_{i.} = n_{i.}/n_{..} \qquad \text{V.1}$$
$$\hat{\theta}_{.j} = n_{.j}/n$$

If the control and complement are independent,

$$\theta_{ij} = \theta_{i.}\theta_{.j} \qquad \text{V.2}$$

These simple assumptions allow for a thorough investigation of the possible interrelationships within the data sets.

Another way to view the assumption of independence is through the use of proportions. Thus, if the basic division is made via the quality of the control, the the proportions

$$P_1 = n_{11}/n_{1.} \quad \text{(respectively } P_2 = n_{21}/n_{2.}) \qquad \text{V.3}$$

represent, in words, the proportion of the years that have quality 1 in the control and have quality 1 (respectively 2) in the complement.

The question of independence may be approached in several ways as described below.

1. <u>Fishers Exact Test</u>

A test for the significance of any dependence was proposed by Fisher in the case in which the marginal totals, $n_{i.}, n_{.j}$, and $n_{..}$ are known a priori (cf. Brownlee [Ref. 6]). To draw from Brownlee, knowledge of the marginals and n_{11} gives knowledge of all the other elements of the table. The

probability of the event of having exactly n_{11} years that display quality 1 in both the complement and the control is;

$$P(N_{11}=n_{11}) = \frac{n_{1.}!n_{.1}!n_{.2}!n_{2.}!}{n_{..}!n_{11}!n_{12}!n_{21}!n_{22}!} \qquad \text{V.4}$$

A test may then be applied, using V.4, to determine the significance of any dependence. This test is usually applied by simply summing these probabilities in the tail of the distribution (V.4) in the same direction as the noted extreme.

The usual procedure to provide a two-sided test of significance is to double a one sided figure. This procedure is acceptable due to the symmetric appearance of the distribution.

Under the assumptions of independence

$$E\left[N_{11}\right] = \frac{n_{1.}n_{.1}}{n_{..}} \qquad \text{V.5}$$

$$V\left[N_{11}\right] = \frac{n_{1.}n_{.1}n_{2.}n_{.2}}{n_{..}\ (n_{..}-1)} \qquad \text{V.6}$$

and the random variable U defined as

$$U = \frac{N_{11} - E[N_{11}]}{\sqrt{V[N_{11}]}} \qquad \text{V.7}$$

is asymptotically distributed as a normal random variable with mean zero and variance one. This asymptotic result combined with a continuity correction yields a test statistic

of

$$u' = \frac{\{|n_{11}\ n_{22} - n_{12}\ n_{21}| - n_{..}/2\}\sqrt{n_{..}}}{\sqrt{n_{1.}n_{.1}n_{2.}n_{.2}}} \qquad V.8$$

The statistic U' may then be used as a test, using standard normal tables, of the significance of any variation from the assumption of independence. It should be noted at this point that if the random variable U' is squared, U'^2 will be distributed as a chi-square with one degree of freedom. The squaring of U' with simplifying algebra yields the Yates correction to a standard chi-squared goodness of fit statistic

$$(u')^2 = \frac{\{|n_{11}\ n_{22} - n_{12}\ n_{21}| - n_{..}/2\}^2\ n_{..}}{(n_{11}+n_{12})(n_{11}+n_{21})(n_{12}+n_{22})(n_{21}+n_{22})} \qquad V.9$$

see Dixon and Massey [Ref. 5]. This allows the use of the chi-square tables as an equivalent test to that of V.8.

2. Odds

Subsection V.A.1 above deals with the significance of any observed interdepedence between the control and the complement. The question of the degree of dependence should also be addressed. The measure to be used is the odds ratio. Using the notation of Fleiss [Ref. 3], a measure of seeing quality 1 in the complement $\underline{Y}$ may be

$$\Omega_1 = \frac{P(Y=1 \mid X=1)}{P(Y=2 \mid X-1)}\ ; \qquad V.10$$

this is then the odds that quality 1 will occur in the complement given that quality 1 is present in the control. In a

similar manner;

$$\Omega_2 = \frac{P(Y=1 \mid X=2)}{P(Y=2 \mid X=2)} \qquad \text{V.11}$$

is the odds that quality 1 will occur in the complement given that quality 2 was observed in the control. The currently most often used measure is the odds ratio ω, or

$$\omega = \frac{\Omega_1}{\Omega_2} \qquad \text{V.12}$$

Note that, if the appearance of quality 1 in the complement is independent of whether or not it appears in the control, then $\omega = 1$. While $\omega > 1$ implies that the odds of the complement having quality 1, given that quality 1 was observed in the control, are greater than the odds of the complement having quality 2. This would indicate that the control would be some sort of predictor for the complement, relative to the selected qualities.

In the same continuity correcting spirit, as was used with the Yates chi-square, an estimate for ω may be obtained from the Table as

$$\hat{\omega} = 0 = \frac{(n_{11}+.5)\ (n_{22}+.5)}{(n_{12}+.5)\ (n_{21}+.5)} \qquad \text{V.13}$$

with a standard error of

$$\text{s.e.}(0) = 0\sqrt{\frac{1}{n_{11}+.5} + \frac{1}{n_{12}+.5} + \frac{1}{n_{21}+.5} + \frac{1}{n_{22}+.5}} \qquad \text{V.14}$$

The natural logarithm of this odds ratio will be discussed more fully in section VI.

B. ANALYSIS

The theory of subsection A above is applied to the three data sets as discussed below. The control is typically taken as a monthly anomaly, say October. Here, the quality is taken as either a positive or a negative anomaly. Thus $X=1$ occurs when the month of October falls below its mean and $X=2$ occurs when it falls above its mean. The complement consists of the sum of the rainfall for the succeeding eleven months, or in symbols:

$$X_t = R_{t,1} - \overline{R}_t.$$

$$Y_t = \sum_{m=2}^{12} R_{t,m} - \frac{1}{N}\sum_{t=1}^{N}\left(\sum_{m=2}^{12} R_{t,m}\right). \qquad \text{V.15}$$

Where it is understood that $X=1$ when $X_t < 0$, $X=2$ when $X_t > 0$ and similarly for Y_t.

Various control subsets are used; October through September were investigated by themselves as were all adjacent pairs, triples, and four-tuples of months. For an example, consider the spring (April, May, and June) and its complement (July through March). In this case

$$X_t = \sum_{m=7}^{9} R_{t,m} - \frac{1}{N}\sum_{t=1}^{N-1}\left(\sum_{m=7}^{9} R_{t,m}\right) \qquad \text{V.16}$$

$$Y_t = \sum_{m=10}^{12} R_{t,m} + \sum_{m=1}^{6} R_{t+1,m} - \frac{1}{N-1}\sum_{t=1}^{N-1}\left(\sum_{m=10}^{12} R_{t,m} + \sum_{m=1}^{6} R_{t+1,m}\right).$$

Equations V.15 and V.16 imply that the data are always analyzed as deviations from the arithmetic mean. However, the data are also analyzed as deviations from the median

and the lower quartile. In the tables to follow, 'A' refers to both control and complement having the arithmetic mean removed, 'M' refers to both control and complement having their respective medians removed, and 'QL' refers to the control having the lower quartile removed while the median was removed from the complement.

The first four Tables (37 through 40), give the significance levels of observed departures from independence of the control and complement. Only those values having a Yates corrected chi-square of greater than 1.00 are listed. The entries represent the two-tailed probability of a random deviation in excess of that observed. Although the cut off criterion was the Yates chi-square, the agreement between its probability and that obtained from the Fisher exact and normal tests did not differ in the first two decimal places.

TABLE 37

SIGNIFICANCE OF OBSERVED DEPARTURES FROM INDEPENDENCE OF SINGLE MONTH CONTROL VERSUS SUCCEEDING ELEVEN MONTH COMPLEMENTS

Data set		RN			FL			SC	
Differentiator	A	M	QL	A	M	QL	A	M	QL
Control									
October	.18	.22	.24			.31		.15	
November					.10	.12			.28
December									
January	.14	.30	.19	.002	.04	.12	.21	.19	
February								.11	
March								.13	
April								.18	
May		.14							
June	.24								
July		.31					.21		
August									
September		.14	.12						

TABLE 38

SIGNIFICANCE OF OBSERVED DEPARTURES FROM INDEPENDENCE OF PAIRS OF MONTHS VERSUS SUCCEEDING TEN MONTH COMPLEMENTS

Data Set	RN			FL			SC		
Differentiators	A	M	QL	A	M	QL	A	M	QL
Control									
Oct+Nov	.18					.09	.24		
Nov+Dec					.10	.12	.15	.31	.08
Jan+Feb	.30	.30		.01	.02	.12	.11	.19	
Feb+Mar									.11
Mar+Apr									
Apr+May									
May+Jun		.14	.12	.24					
Jun+Jul									
Jul+Aug			.19		.18				
Aug+Sep									

TABLE 39

SIGNIFICANCE OF OBSERVED DEPARTURES FROM INDEPENDENCE OF TRIPLES OF MONTHS VERSUS SUCCEEDING NINE-MONTH COMPLEMENTS

Data Set	RN			FL			SC		
Differentiators	A	M	QL	A	M	QL	A	M	QL
Control									
Oct+Nov+Dec							.19		
Nov+Dec+Jan									
Dec+Jan+Feb				.31					
Jan+Feb+Mar									
Feb+Mar+Apr								.31	.28
Mar+Apr+May							.22		
Apr+May+Jun			.30						
May+Jun+Jul		.14	.30						
Jun+Jul+Aug					.18				
Jul+Aug+Sep									

TABLE 40

SIGNIFICANCE OF OBSERVED DEPARTURES FROM INDEPENDENCE OF FOUR-TUPLES OF MONTHS VERSUS SUCCEEDING EIGHT-MONTH COMPLEMENTS

Data set	RN			FL			SC		
Differentiator	A	M	QL	A	M	QL	A	M	QL
Control									
Oct+Nov+Dec+Jan				.26					
Nov+Dec+Jan+Feb		.22							
Dec+Jan+Feb+Mar		.22							
Jan+Feb+Mar+Apr								.31	
Feb+Mar+Apr+May	.22						.27		
Mar+Apr+May+Jun							.20	.31	
Apr+May+Jun+Jul			.30						.28
May+Jun+Jul+Aug		.14							
Jun+Jul+Aug+Sep									

Several choices for predictors are suggested in the previous tables. However, the apparent strongest candidate for a predictor is January. The control of January by itself and January paired with December, are the most consistently significant entries. Tables 41 below gives the odds ratio, V.13, for January, and January and December, as controls.

TABLE 41

ODDS RATIO OF JANUARY VERSUS FEBRUARY THROUGH DECEMBER AND JANUARY PLUS DECEMBER VERSUS FEBRUARY THROUGH NOVEMBER

Differentiator	A	M	QL
Data set RN			
January	4.59	3.15	5.13
Jan+Dec	3.15	3.15	2.01
Data set FL			
January	10.71	4.72	4.30
Jan+Dec	7.42	6.02	4.30
Data set SC			
January	2.45	2.49	2.23
Jan+Dec	3.00	2.49	1.44

At this point in the analysis it was decided to explore more fully the power of January as a predictor. It should be stated that other possibilities for predictors are suggested by the tables, but time did not allow an exhaustive study of all of these.

C. OTHER RESULTS

The results of section V.B suggest that a more detailed analysis of January as a predictor is in order. The first method tried for this was ordinary least squares regression of the rainfall total in January versus the total for February through December. This is the model below.

Let $X_t = R_{t,4}$

$$Y_t = \sum_{m=5}^{12} R_{t,m} + \sum_{m=1}^{3} R_{t+1,m} \qquad \text{V.17}$$

then assume that

$$Y_t = \alpha + \beta X_t + \varepsilon_t \qquad \text{V.18}$$

as the standard, linear model where $\{\varepsilon_t\}$ are assumed to be independent and identically distributed with mean zero and variance $\sigma\varepsilon^2$. If the predictability of January is strong, this model, V.18, may result in a good fit of the data. Table 42 below is the resulting ANOVA for this regression. As may easily be seen, the model does not appear to have any significance.

TABLE 42

ANOVA FOR REGRESSION OF SIMPLE
LINEAR MODEL FOR ALL DATA SETS

R-squared = .1033
Standard error of estimate = 4.1574 RN

SOURCE	DF	SS	MS	F
Total	22	404.798		
Regression	1	41.821	41.821	2.42
Jan-Control	1	41.821	41.821	2.42
Residual	21	362.977	17.285	

Variable	Coefficient	Standard error	T
Alpha	.2178	.878	.25
Beta	.5993	.385	1.56

	90% Confidence Limits Lower limit	Upper limit
Alpha	-1.1817	1.6174
Beta	- .0148	1.2133

R-squared = .2066
Standard error of estimate = 4.2804 FL

AOV

SOURCE	DF	SS	MS	F
Total	35	785.196		
Regression	1	162.252	162.252	8.86
Jan-Control	1	162.252	162.252	8.86
Residual	34	622.945	18.322	

Variable	Coefficient	Standard error	T
Alpha	-.2670	.715	-.37
Beta	1.0095	.339	2.98

	90% Confidence Limits	
	Lower limit	Upper limit
Alpha	- .2670	.8543
Beta	1.0095	1.5417

R-squared = .0426
Standard error of estimate = 6.1508 SC

AOV

SOURCE	DF	SS	MS	F
Total	46	1778.131		
Regression	1	75.687	75.687	2.00
Jan-Control	1	75.687	75.687	2.00
Residual	45	1702.444	37.832	

Variable	Coefficient	Standard error	T
Alpha	-.2771	.898	-.31
Beta	.4355	.308	1.41

	90% Confidence Limits	
	Lower limit	Upper limit
Alpha	-1.6766	1.1224
Beta	- .0446	.9157

The same model with means removed, V.19, below was tried and, although slightly better, is still not strong.

$$Y_t - \bar{Y}. = \alpha + \beta(X_t - \bar{X}.) + \varepsilon_t \qquad \text{V.19}$$

TABLE 43

ANOVA FOR REGRESSION OF SIMPLE LINEAR
MODEL WITH MEANS REMOVED FOR ALL DATA SETS

R-squared = .1033
Standard error of estimate = 4.1575 RN

AOV

SOURCE	DF	SS	MS	F
Total	22	404.798		
Regression	1	41.821	41.821	2.42
Jan-Control	1	41.821	41.821	2.42
Residual	21	362.977	17.285	

Variable	Coefficient	Standard error	T
Alpha	11.7215	1.851	6.33
Beta	.5993	.385	1.56

90% Confidence Limits

	Lower limit	Upper limit
Alpha	8.7236	13.2483
	-.0148	1.2133

R-squared = .1747
Standard error of estimate = 4.4276 FL

AOV

SOURCE	DF	SS	MS	F
Total	35	807.618		
Regression	1	141.108	141.108	7.20
Jan-Control	1	141.108	141.108	7.20
Residual	34	666.510	19.603	

Variable	Coefficient	Standard error	T
Alpha	10.9860	1.442	7.62
Beta	.9414	.351	2.68

90% Confidence Limits

	Lower limit	Upper limit
Alpha	8.7236	13.2483
Beta	.3909	1.4920

R-squared = .5617 SC
Standard error of estimate = .7589

AOV

SOURCE	DF	SS	MS	F
Total	7	7.884		
Regression	1	4.428	4.428	7.69
Jan-Control	1	4.428	4.428	7.69
Residual	6	3.456	.576	

Variable	Coefficient	Standard error	T
Alpha	- .1946	.287	-.68
Beta	1.7710	.637	2.77

90% Confidence Limits

	Lower limit	Upper limit
Alpha	-.7040	.3148
Beta	.6362	2.9058

The amount of rainfall in January does not appear to be a strong predictor for the amount of rainfall in February through December. This seems to indicate that the relationships between January rainfall and rainfall during the next eleven months is not as strong as expected. However, a further technique is available, that of log-odds and logistic regression, which are the subjects of the next section.

VI. LOGISTIC ANALYSIS

A. THEORY

The logistic analysis to be described in the section was developed from Gaver [Ref. 2] and Fleiss [Ref. 3]. This analysis derives from the model of I.5 as stated in the introduction.

The basic approach is to view the complement as having a binary representation, with success being defined as a complement above its mean (see equations I.3 and I.4) and failure as the complement below its mean. The problem then is to find the conditional probability of a success (the complement being above its mean for a year) given that the control (January rainfall) takes on a particular value.

The control is now taken to be the logged rainfall anomaly of January and is found in equation I.2 repeated below.

$$X_t = \ell n(R_{t,4}) - \frac{1}{N} \sum_{t=1}^{N} \ell n(R_{t,4}) \qquad \text{VI.1}$$

If the probability of success, given X_t is written as

$$P(\text{Success}|X_t) = \theta_t \qquad \text{VI.2}$$

a superficially attractive model for θ_t is

$$\theta_t = \alpha + \beta X_t + \varepsilon_t \qquad \text{VI.3}$$

This model has two difficulties, the worst of which is that probabilities of greater than one or less that zero are

allowed. Secondly, the θ_t are not available in proportion form to fit the model.

Initially, the problem of estimating θ_t is approached by grouping the data. Section II indicated that each year seemed to be independent of the next and that there was no trend. This allows for the ordering of the X_t's into their order statistics $X_{(t)}$'s. Once the ordering has been done, non overlapping groups of arbitrary size may be formed as shown. Let $X_1, X_2, X_3, \ldots, X_{12}$ be a series of 12 years from an arbitrary data set, with associated order astatistics $X_{(1)}, X_{(2)}, X_{(3)}, \ldots, X_{(12)}$. The if groups of size three are desired, the data are partitioned below.

$$X_{(1)},\ X_{(2)}, X_{(3)} | X_{(4)}, X_{(5)}, X_{(6)} | \ldots | X_{(10)}, X_{(11)}, X_{(12)}$$

Given these groups, let $\tilde{X}_j$ be a measure of location for the j^{th} group. This analysis used the median, therefore $\tilde{X}_j = X_{(3j+1)}$. Also associated with each group is R_j (not rainfall), the number of success in group j, and n_j, the number of elements in group j. From this set up, the required probabilities may be estimated as;

$$\hat{\theta}_j = R_j/n_j \qquad \text{VI.4}$$

A solution to the first problem, that of the model yielding probabilities outside of (0,1), is to use the log odds, instead of θ_j where;

$$\text{Log odds} = \phi_j = \ell n\left(\frac{\theta_j}{1-\theta_j}\right) \qquad \text{VI.5}$$

which is equivalent to the logarithm of the odds ratio as given in V.13. Gaver [Ref. 2] suggests that a correction of .5 be applied to guard against the problem of 0 and 1 within the logarithm and to reduce the bias. The statistic then becomes;

$$\phi_j' = \ln\left(\frac{\theta_j+.5}{1.5-\theta_j}\right) \qquad \text{VI.6}$$

The temptation is to go directly to the model

$$\phi_j' = \alpha + \beta \quad \tilde{X}_j; \qquad \text{V1.7}$$

yet $V[\theta_j] = \theta_j(1-\theta_j)/n_j$ which is not constant. This suggests the need for a weighting scheme.

The weighting scheme used was that of iteratively reweighted least squares, (IRWLS), using the bi-weights. This method is dicussed in detail in Mosteller and Tukey [Ref. 13].

Although grouping of the data and the model of VI.7 provide adequate representation of the underlying structure, the logistic model itself, I.5, when viewed through the eyes of maximum likelihood theory can yield more insight.

The model is assumed to be

$$\theta_t = \frac{e^{\alpha+\beta X_t}}{1+ e^{\alpha-\beta X_t}} \qquad \text{VI.8}$$

where the X_t are independent.

The likelihood function is then

$$L(X,Y;\alpha,\beta) = \prod_{t=1}^{N} \left(\frac{e^{\alpha+\beta X_t}}{1 + e^{\alpha+\beta X_t}}\right)^{Y_t} \left(\frac{1}{1 + e^{\alpha+\beta X_t}}\right)^{1-Y_t}$$

$$= \prod_{t=1}^{N} \frac{e^{\alpha Y_t + \beta X_t Y_t}}{\prod_{t=1}^{N} (1+e^{\alpha+\beta X_t})} \qquad \text{V.9}$$

and the log-likelihood is

$$\mathcal{L}(X,Y;\alpha,\beta) = \alpha \sum_{t=1}^{N} Y_t + \beta \sum_{t=1}^{N} X_t Y_t - \sum_{t=1}^{N} \ln (1 + e^{\alpha+\beta X_t}) \qquad \text{V.10}$$

The gradient, and Hessian of VI.9 are

$$\frac{\partial \mathcal{L}}{\partial \alpha} = \sum_{t=1}^{N} Y_t - \sum_{t=1}^{N} \left(\frac{\psi_t}{1 + \psi_t}\right)$$

$$\frac{\partial \mathcal{L}}{\partial \beta} = \sum_{t=1}^{N} X_t Y_t - \sum_{t=1}^{N} X_t \left(\frac{\psi_t}{1 + \psi_t}\right) \qquad \text{V.11}$$

and

$$H_{\mathcal{L}} = \begin{pmatrix} -\sum_{t=1}^{N} \frac{\psi_t}{(1+\psi_t)^2} & -\sum_{t=1}^{N} \frac{X_t \psi_t}{(1+\psi_t)^2} \\ -\sum_{t=1}^{N} \frac{X_t \psi_t}{(1+\psi_t)^2} & -\sum_{t=1}^{N} \frac{X_t^2 \psi_t}{(1+\psi_t)^2} \end{pmatrix}$$

where $\psi_t = e^{\alpha+\beta X_t}$.

A simple way to solve for $\hat{\alpha}$ and $\hat{\beta}$ is to use Newtons Method as,

$$\begin{pmatrix}\alpha \\ \beta\end{pmatrix}_{k+1} = \begin{pmatrix}\alpha \\ \beta\end{pmatrix}_{k} - H_L^{-1} \nabla {}^{t}L$$

since all necessary elements may be calculated in one pass of the computor algorithm.

One beneficial byproduct of the maximum likelihood approach is the asymptotic information matrix, H_L^{-1}. Gaver [Ref. 2] states that the diagonal elements of this matrix provide good estimates of $V[\hat{\alpha}]$ and $V[\hat{\beta}]$ under assumptions of normality.

B. ANALYSIS

1. Grouped Data

The first approach taken was that of grouping the data as described above. Groups of 3, 4, and 5 were used, as were two separate methods of regression, ordinary least squares (OLS), and iteratively reweighted least squares (IRWLS). Tables 44 (RN), 45 (FL), and 46 (SC) present the data and Tables 47 (OLS), and 48 (IRWLS) present the results of the regressions.

TABLE 44

DATA SET RN, LOGGED JANUARY ANOMALIES AND SUCCESSES FOR GROUPED AND UNGROUPED FORMS

UNGROUPED			GROUPED		
t	X	Y	j	X	Y
1	-1.22	0		group=3	
2	-1.09	0	1	-1.09	0
3	- .66	0	2	- .46	1
4	- .58	0	3	- .19	1
5	- .46	0	4	.01	2
6	- .36	1	5	.16	2
7	- .31	0	6	.36	2
8	- .18	1	7	.50	2
9	- .17	0	8	.98	2
10	- .05	1		group=4	
11	.01	1	1	-.38	0
12	.01	0	2	-.34	2
13	.15	1	3	-.03	2
14	.16	0	4	.20	2
15	.24	1	5	.47	4
16	.28	0	6	.94	2
17	.36	1		group=5	
18	.46	1	1	-.66	0
19	.48	1	2	-.19	3
20	.50	1	3	.15	3
21	.51	0	4	.46	4
22	.94	1	5	.94	2
23	1.01				

TABLE 45

DATA SET FL, LOGGED JANUARY ANOMALIES AND SUCCESSES FOR GROUPED AND UNGROUPED DATA

UNGROUPED			GROUPED		
t	X	Y	j	X	Y
1	-3.41	1		group=3	
2	-1.82	0	1	-1.82	1
3	-1.14	0	2	-1.00	0
4	-1.02	0	3	- .46	0
5	-1.00	0	4	- .20	2
6	- .92	0	5	.11	1
7	- .47	0	6	.14	1
8	- .46	0	7	.20	0
9	- .39	0	8	.34	2
10	- .28	1	9	.50	2
11	- .20	0	10	.61	2
12	- .14	1	11	.75	3
13	- .04	1	12	1.14	3
14	.11	0		group=4	
15	.11	0	1	-1.48	1
16	.13	0	2	- .70	0
17	.14	0	3	- .24	2
18	.16	1	4	.11	1
19	.16	0	5	.16	1
20	.20	0	6	.32	2
21	.25	0	7	.50	2
22	.30	1	8	.71	4
23	.34	1	9	1.11	4
24	.38	0		group=5	
25	.46	1	1	-1.14	1
26	.50	0	2	- .46	1
27	.51	1	3	- .04	2
28	.57	0	4	.16	1
29	.61	1	5	.34	3
30	.71	1	6	.57	3
31	.71	1	7	.95	6
32	.75	1			
33	.82	1			
34	1.08	1			
35	1.14	1			
36	1.31	1			

TABLE 46

DATA SET SC, LOGGED JANUARY ANOMALIES AND SUCCESSES FOR GROUPED AND UNGROUPED FORMS

UNGROUPED

t	X	Y
1	-4.24	1
2	-1.41	0
3	-1.38	0
4	-1.08	0
5	-1.07	0
6	-1.06	0
7	-1.02	0
8	- .96	1
9	- .65	0
10	- .60	0
11	- .55	1
12	- .36	0
13	- .27	1
14	- .26	0
15	- .21	1
16	- .17	1
17	- .10	0
18	- .03	0
19	- .01	1
20	.02	0
21	.12	1
22	.14	0
23	.20	0
24	.22	0
25	.23	0
26	.28	1
27	.28	1
28	.40	1
29	.40	1
30	.41	0
31	.43	0
32	.46	1
33	.49	0
34	.56	1
35	.56	1
36	.60	0
37	.64	0
38	.69	0
39	.70	0
40	.70	1
41	.78	1
42	.81	0
43	.88	0
44	.97	0
45	.97	1
46	1.03	1
47	1.47	1

GROUPED

j	X	Y
	group=3	
1	-1.41	1
2	-1.07	0
3	- .96	1
4	- .55	1
5	- .26	2
6	- .10	1
7	.02	2
8	.20	0
9	.28	2
10	.40	2
11	.46	1
12	.56	2
13	.69	0
14	.78	2
15	.97	1
16	1.25	2
	group=4	
1	-1.39	1
2	-1.04	1
3	- .58	1
4	.23	3
5	- .02	1
6	.17	1
7	.28	3
8	.42	2
9	.56	2
10	.69	1
11	.78	1
12	1.03	3
	group=5	
1	-1.38	1
2	- .96	1
3	- .27	3
4	- .04	2
5	.20	1
6	.40	4
7	.49	3
8	.69	1
9	.88	2
10	1.25	2

TABLE 47a

ORDINARY LEAST SQUARES REGRESSION WITH
THE MODEL OF EQUATION VI.7 FOR DATA SET RN

R-squared = .0423
Standard error of estimate = 6.1508

RN
GROUP=3

AOV

SOURCE	DF	SS	MS	F
Total	46	1778.131		
Regression	1	75.687	75.687	2.00
Jan-Control	1	75.687	75.687	2.00
Residual	45	1702.444	37.832	

Variable	Coefficient	Standard error	T
Alpha	14.5833	1.677	8.69
Beta	.4355	.308	1.41

90% Confidence Limits

	Lower limit	Upper limit
Alpha	11.9680	17.1988
Beta	- .0446	.9157

R-squared = .4068
Standard error of estimate = 1.2099

RN
GROUP=4

AOV

SOURCE	DF	SS	MS	F
Total	5	9.871		
Regression	1	4.016	4.016	2.74
Jan-Control	1	4.016	4.016	2.74
Residual	4	5.855	1.464	

Variable	Coefficient	Standard error	T
Alpha	-.1696	.517	-.33
Beta	1.7687	1.068	1.66

90% Confidence Limits

	Lower limit	Upper limit
Alpha	-1.1675	.8284
Beta	1.7688	3.8288

R-squared = .5973 RN
Standard error of estimate = .9995 GROUP=5

AOV

SOURCE	DF	SS	MS	F
Total	4	7.442		
Regression	1	4.446	4.446	4.45
Jan-Control	1	4.446	4.446	4.45
Residual	3	2.997	.999	

Variable	Coefficient	Standard error	T
Alpha	-.2647	.461	-.57
Beta	1.7193	.815	2.11

90% Confidence Limits

	Lower limit	Upper limit
Alpha	-1.2358	.7064
Beta	.0040	3.4345

TABLE 47b

ORDINARY LEAST SQUARES REGRESSION WITH
THE MODEL OF EQUATION VI.7 FOR DATA SET FL

R-squared = .2862 FL
Standard error of estimate = .8949 GROUP=3

AOV

SOURCE	DF	SS	MS	F
Total	11	20.591		
Regression	1	7.998	7.998	6.35
Jan-Control	1	7.998	7.998	6.35
Residual	10	12.593	1.259	

Variable	Coefficient	Standard error	T
Alpha	-.1459	.324	-.33
Beta	1.0521	.417	2.52

90% Confidence Limits

	Lower limit	Upper limit
Alpha	-.6871	.4944
Beta	.3550	1.7492

R-squared = .5370 FL
Standard error of estimate = 1.0484 GROUP=4

AOV

SOURCE	DF	SS	MS	F
Total	8	16.615		
Regression	1	8.922	8.922	8.12
Jan-Control	1	8.922	8.922	8.12
Residual	7	7.963	1.099	

Variable	Coefficient	Standard Error	T
Alpha	-.1148	.350	-.33
Beta	1.3572	.476	2.83

90% Confidence Limits

	Lower limit	Upper limit
Alpha	-.7237	.4941
Beta	.5297	2.1847

R-squared = .6311 FL
Standard error of estimate = .8809 GROUP=5

AOV

SOURCE	DF	SS	MS	F
Total	6	10.519		
Regression	1	6.639	6.639	8.56
Jan-Control	1	6.639	6.639	8.56
Residual	5	3.880	.776	

Variable	Coefficient	Standard Error	T
Alpha	-.1393	.334	-.42
Beta	1.5225	.521	2.92

90% Confidence Limits

	Lower limit	Upper limit
Alpha	-.7525	.4740
Beta	.5673	2.4777

TABLE 47c

ORDINARY LEAST SQUARES REGRESSION WITH
THE MODEL OF EQUATION VI.7 FOR DATA SET SC

R-squared = .1404 SC
Standard error of estimate - .9901 GROUP=3

AOV

SOURCE	DF	SS	MS	F
Total	15	15.965		
Regression	1	2.241	2.241	2.29
Jan-Control	1	2.241	2.241	2.29
Residual	14	13.724	.980	

Variable	Coefficient	Standard Error	T
Alpha	-.3039	.249	-1.22
Beta	.5059	.335	1.51

	90% Confidence Limits	
	Lower limit	Upper limit
Alpha	-.7090	.1012
Beta	-.0386	1.0504

R-squared = .1872 SC
Standard error of estimate = .8959 GROUP=4

AOV

SOURCE	DF	SS	MS	F
Total	11	9.875		
Regression	1	1.848	1.848	2.30
Jan-Control	1	1.848	1.848	2.30
Residual	10	8.026	.803	

Variable	Coefficient	Standard Error	T
Alpha	-.2284	.260	-.88
Beta	.5447	.359	1.52

	90% Confidence Limits	
	Lower limit	Upper limit
Alpha	-.6618	.2051
Beta	.5447	1.1439

R=squared = .2862 SC
Standard error of estimate = .8949 GROUP=5

ADV

SOURCE	DF	SS	MS	F
Total	9	8.976		
Regression	1	2.569	2.569	3.21
Jan-Control	1	2.569	2.569	3.21
Residual	8	6.407	.801	

Variable	Coefficient	Standard Error	T
Alpha	-.1838	.287	-.64
Beta	.6573	.367	1.79

90% Confidence Limits

	Lower limit	Upper limit
Alpha	-.6736	.3061
Beta	.0303	1.2843

TABLE 48

ITERATIVELY REWEIGHTED LEAST SQUARES
REGRESSION USING BI-WEIGHTS FOR THE MODEL OF EQUATION VI.7

		$\hat{\alpha}$	$\hat{\beta}$
		GROUP SIZE 3	
C=9			
Data sets	RN	.048	1.013
	FL	-.063	.664
	SC	-.116	.351
C=4			
Data sets	RN	.052	1.031
	FL	-.197	1.049
	SC	.154	.480
		GROUP SIZE 4	
C=9			
Data sets	RN	-.103	.665
	FL	-.093	.705
	SC	-.169	.280
		GROUP SIZE 5	
C=9			
Data sets	RN	-.065	.865
	FL	-.107	.723
	SC	-.126	.413

2. Maximum Likelihood

Table 49 displays the final points for each data set along with the inverse Hessian at that point. Figures 103 (RN), 104 (FL), and 105 (SC) are interesting in that they portray the contours of the likelihood functions for each data set. These contours show the variance of the estimated parameters in a graphic way. Note how data sets RN and FL seem to have some sort of horizontal ridge indicating a good pick of the slope parameter, yet the contours of data set SC are almost circular about the origin of the axes indicating no significant difference from zero for either parameter.

TABLE 49

MAXIMUM LIKELIHOOD ESTIMATES OF $\hat{\alpha}$ AND $\hat{\beta}$ ALONG WITH ESTIMATES OF THEIR VARIANCE FOR ALL THREE DATA SETS

Data set RN

$$\hat{\alpha} = .062 \qquad \hat{\beta} = 2.918 \qquad H_L^{-1} = \begin{pmatrix} .257 & -.043 \\ -.034 & 1.720 \end{pmatrix}$$

$$V[\hat{\alpha}] = .257$$
$$V[\hat{\beta}] = 1.720$$

Data set FL

$$\hat{\alpha} = -.171 \qquad \hat{\beta} = .933 \qquad H_L^{-1} = \begin{pmatrix} .129 & -.040 \\ -.040 & .298 \end{pmatrix}$$

$$V[\hat{\alpha}] = .129$$
$$V[\hat{\beta}] = .298$$

Data set SC

$$\hat{\alpha} = -.303 \qquad \hat{\beta} = .171 \qquad H_L^{-1} = \begin{pmatrix} .088 & -.035 \\ -.035 & .113 \end{pmatrix}$$

$$V[\hat{\alpha}] = .088$$
$$V[\hat{\beta}] = .113$$

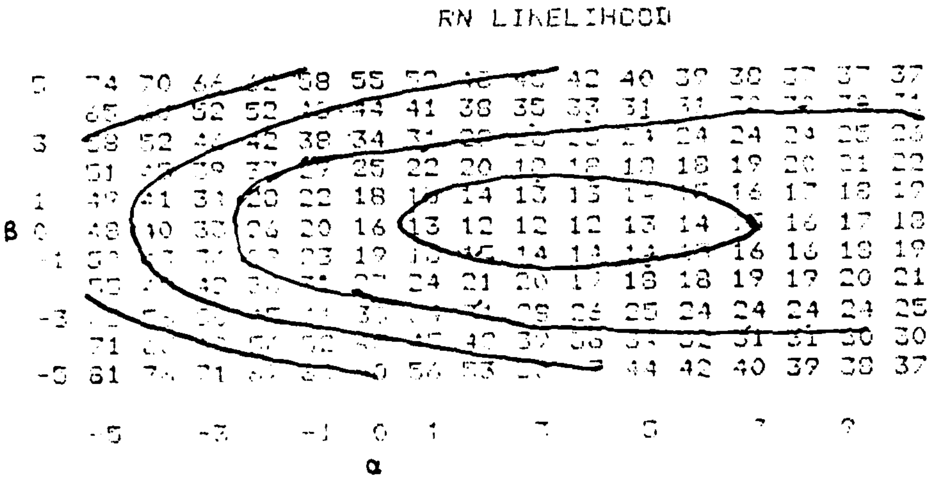


Figure 103. Contours of log likelihood function for data set RN

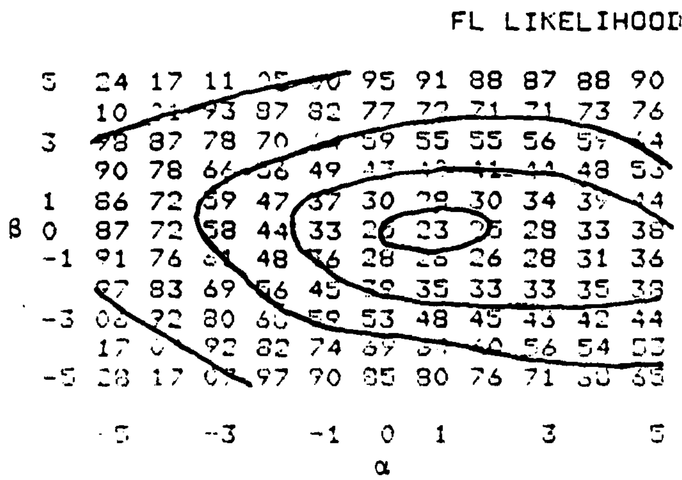


Figure 104. Contours of log likelihood function for data set FL

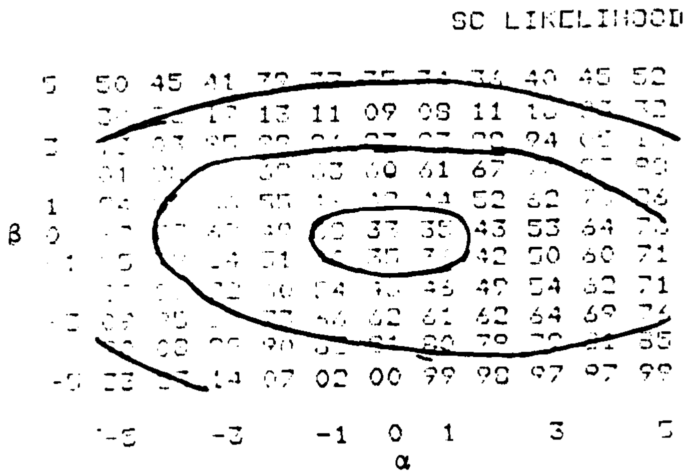


Figure 105. Contours of log likelihood function for data set SC

C. DISCUSSION

A recapitulation of all the parameter values is in Table 50. Some interesting observations to be made from this table are:

(1) The slope for data set RN, as found by maximum likelihood, is much greater than that of any other method or data set. The first temptation is to treat this as an outlier, yet the evidence of the contour plot and of the validation of the next section tend to back up this number. The reason for this difference is a possible subject for further reserach.

(2) Except for the intercept values, the slopes of data sets RN and FL seem to be fairly consistent within and between

regression methods. This comment is made in light of the difference of these data sets and that of SC.

(3) Data sets RN and FL seem to be similar in many ways, yet data set SC appears to be different in both degree and significance.

No other strong pattern is apparent in these parameter values. Graphical displays of the parameters, as used with grouped data are given in Figures 106 (RN), 107 (FL), and 108 (SC). Again, note the significant difference of the maximum likelihood line for data set RN. Table 51 contains the data points from which Figures 106, 107 and 108 were drawn.

After viewing these figures, the maximum likelihood approach is the preferred method for the Peninsula data sets, whereas the robust(C=4) IRWLS may be best for the Valley data set. Although fits were made to data set SC, it appears as if no great significance has been found.

TABLE 50

PARAMETER FIT RECAPITULATION
FOR ALL DATA SETS

Method		RN	FL	SC
			DATA SETS	
1. Maximum	α :	.062	-.171	- .303
Likelihood	β :	2.918	.933	.171
2. OLS				
Group=3	α :	- .195	-.146	- .304
	β :	1.771	1.052	.506
Group=4	α :	- .170	-.115	- .228
	β :	1.769	1.357	.545
Group=5	α :	- .265	-.139	- .184
	β :	1.719	1.523	.657
3. IRWLS				
Group=3	α :	.048	-.063	- .116
C=9	β :	1.013	.664	.351
Group=3	α :	.052	-.197	.154
C=4	β :	1.031	1.049	.480
Group 4	α :	-.103	-.093	- .169
C=9	β :	.665	.705	.280
Group=5	α :	-.065	-.107	- .126
C=9	β :	.855	.723	.413

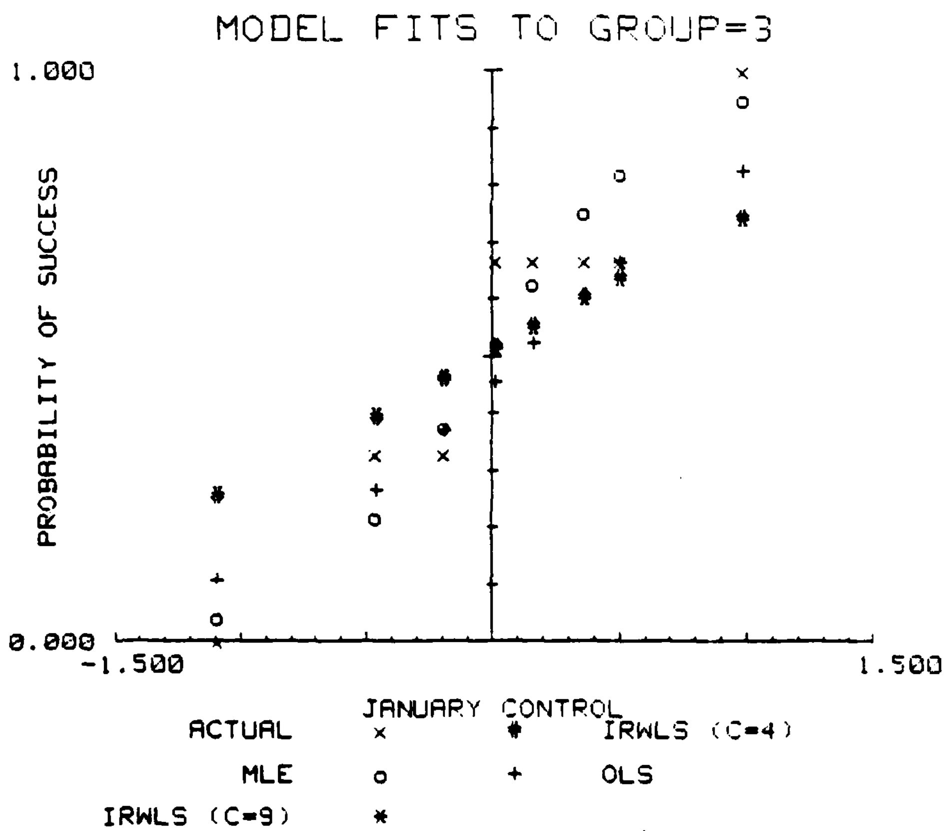


Figure 106. Estimated probability of greater-than-average total rest-of-year rainfall versus the anomaly of logged rainfall for January for data set RN

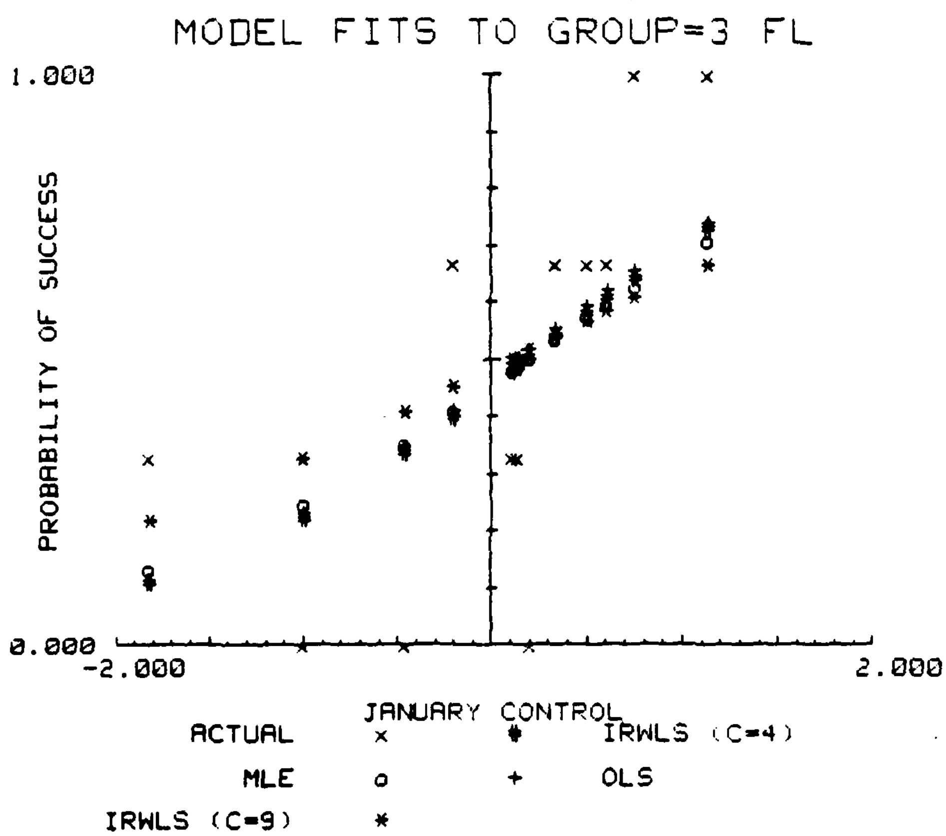


Figure 107. Estimated probability of greater-than-average total rest-of-year rainfall versus the anomaly of logged rainfall for January for data set FL

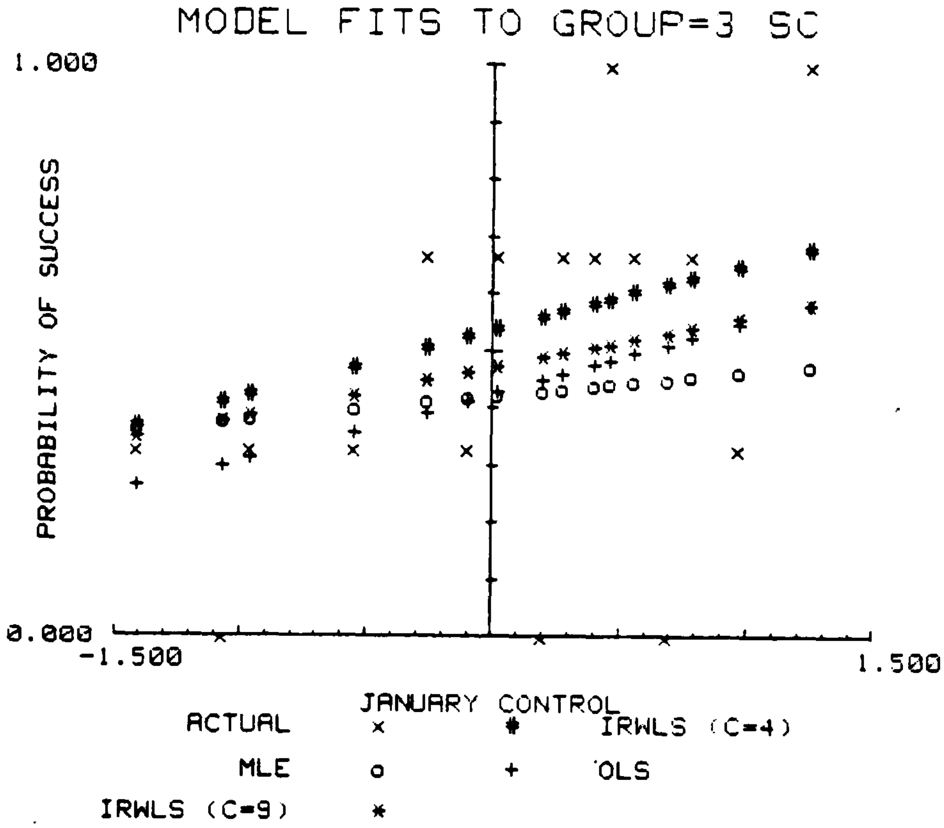


Figure 108. Estimated probability of greater-than average total rest-of-year rainfall versus the anomaly of logged rainfall for January for data set FL

TABLE 51a

ACTUAL VALUES FOR MODEL FITS OF RN

LOGGED ANOMALY	ACTUAL VALUE	MLE	IRWLS C=9	IRWLS C=4	OLS
-1.090	0.000	.042	.258	.255	.107
- .460	.330	.218	.397	.396	.267
- .190	.330	.379	.464	.464	.370
.010	.670	.523	.515	.516	.456
.160	.670	.629	.552	.554	.522
.360	.670	.753	.602	.604	.609
.500	.670	.821	.635	.638	.666
.980	1.000	.949	.739	.743	.824

TABLE 51b

ACTUAL VALUES FOR MODEL FITS OF FL

LOGGED ANOMALY	ACTUAL VALUE	MLE	IRWLS C=9	IRWLS C=4	OLS
-1.820	.330	.124	.219	.108	.113
-1.000	0.000	.249	.326	.223	.232
- .460	0.000	.354	.409	.336	.348
- .200	.670	.412	.451	.400	.412
.110	.330	.483	.503	.480	.492
.140	.330	.490	.507	.487	.500
.200	0.000	.504	.517	.503	.516
.340	.670	.536	.541	.540	.553
.500	.670	.573	.567	.581	.594
.610	.670	.598	.585	.609	.621
.750	1.000	.629	.607	.643	.655
1.140	1.000	.709	.667	.731	.741

TABLE 51c

ACTUAL VALUES FOR MODEL FITS OF SC

LOGGED ANOMALY	ACTUAL VALUE	MEL	IRWLS C=9	IRWLS C=4	OLS
-1.410	.330	.367	.352	.372	.266
-1.070	0.000	.381	.380	.411	.300
- .960	.330	.385	.389	.424	.312
- .550	.330	.402	.423	.473	.358
- .260	.670	.414	.448	.507	.393
- .100	.330	.421	.462	.526	.412
.020	.670	.426	.473	.541	.427
.200	0.000	.433	.489	.562	.449
.280	.670	.437	.496	.572	.460
.400	.670	.442	.506	.586	.475
.460	1.000	.444	.511	.593	.482
.560	.670	.448	.520	.604	.495
.690	0.000	.454	.532	.619	.511
.780	.670	.458	.539	.629	.523
.970	.330	.466	.556	.650	.547
1.250	1.000	.478	.580	.680	.581

VII. VALIDATION OF LOGISTIC MODELS

A. GENERAL

The various parameters that were estimated in the previous section all may be subject to some sort of validation. However, this paper will only view the validation for the maximum likelihood approach on all data sets and the IRWLS (C=4) approach on data set SC. The validation will be conducted against the reserved, independent, data sets of years 1975 through 1980. These are the same data sets as used in section IV.

Table 52 portrays the reserved data, in a form for logistic analysis, and Figure 109 is a display of the derived contingency tables for the reserved data.

TABLE 52

RESERVED DATA IN FORM FOR THE LOGISTIC ANALYSIS

DATA SET RN

YEAR	X	COMPLEMENT	Y_t
1975	-.739	12.92	0
1976	-2.446	11.28	0
1977	-.477	11.14	0
1978	.883	19.62	1
1979	.542	18.11	1
1980	.752	---	-

DATA SET FL

1975	-.470	12.58	0
1976	-2.610	10.58	0
1977	-.470	11.10	0
1978	.734	19.32	1
1979	.547	13.72	0
1980	.633	---	-

DATA SET SC

1975	-.535	17.97	1
1976	-3.773	11.19	0
1977	.352	10.03	0
1978	1.054	23.38	1
1979	.512	16.36	0
1980	.405	---	-

Data set RN

Complement

Control		−	+	
	−	3	0	3
	+	0	2	3
		3	2	5

Data set FL

Complement

Control		−	+	
	−	3	0	3
	+	1	1	2
		4	1	5

Data set SC

Complement

Control		−	+	
	−	1	1	2
	+	2	1	3
		3	2	5

Figure 109. 2x2 contingency Tables of reserved data controlled by the anomaly of January rainfall. The complement is the anomaly of the rest-of-year rainfall.

B. RESULTS

1. Data set RN

The model proposed by the maximum likelihood parameters is

$$\theta_t = \frac{e^{.0618+2.9183X_t}}{1 + e^{.0618+2.9183X_t}} \qquad \text{VIII.1}$$

This model, when applied to the reserved data yields Table 53.

TABLE 53

RESULTS OF LOGSTIC VALIDATION
ON DATA SET RN

YEAR	X	Y	0
1975	-.739	0	.11
1976	-2.746	0	.00
1977	-.477	0	.21
1978	.883	1	.93
1979	.542	1	.84
1980	.752	-	.91

The θ_t is interpreted, again, as: The conditional probability that the complement, the total rainfall for February through December, will be above its mean value, given that the logged January anomaly was X_t. Thus it appears that this model tends to predict the direction of the complements deviation well. Figure 110 is a plot of the estimated probabilities against the actual complement anomaly. For an acceptable fit, this plot should show an upward to the right slope, which it does.

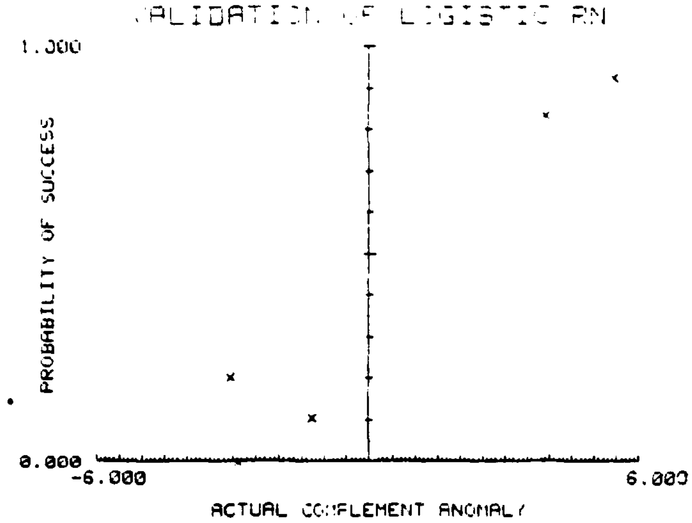


Figure 110. Plot of θ_t versus complement anomalies for data set RN

2. <u>Data Set FL</u>

This data set is quite similar to the RN data, except that the slope parameter is only a third of that of RN. The model is

$$\theta_t = \frac{e^{-.171+.9325X_t}}{1 + e^{-.171+.9325X_t}} \qquad \text{VII.2}$$

TABLE 54

RESULTS OF LOGISTIC VALIDATION
ON DATA SET FL

Year	X	Y	θ
1975	-.470	0	.35
1976	-2.610	0	.07
1977	-.470	0	.35
1978	.734	1	.63
1979	.547	0	.58
1980	.633	-	.60

A plot of the probabilities against the complement anomalies is in Figure 111.

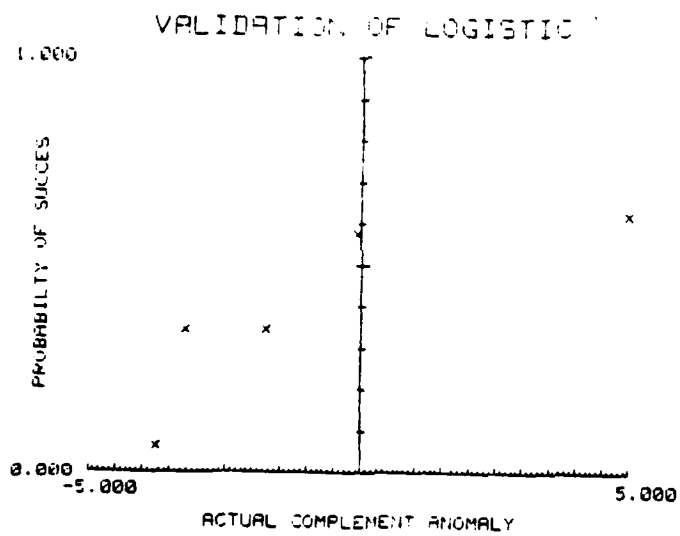


Figure 111. Plot of θ_t versus complement anomalies for data set FL

This fit is not as good as that for data set RN. The outlier, or false prediction of 1979 may not, however, be far out of line. The sparsity of points for which the complement anomaly was positive detracts from the validation effort.

3. <u>Data Set SC</u>

The maximum likelihood model is

$$\theta_t^{(1)} = \frac{e^{-.3025+.171X_t}}{1 + e^{-.3025+.171X_t}} \qquad \text{VII.3}$$

and the IRWLS model is

$$\theta_t^{(2)} = \frac{e^{.1537+.4799X_t}}{1 + e^{.1537+.4799X_t}} \qquad \text{VII.4}$$

and the tabular results are in Table 55.

TABLE 55

RESULTS OF VALIDATION ON DATA SET SC

Year	X	Y	$\theta_t^{(1)}$	$\theta_t^{(2)}$
1975	-.535	1	.40	.47
1976	-3.773	0	.28	.16
1977	.352	0	.44	.58
1978	1.054	1	.47	.66
1979	.512	0	.45	.60
1980	.405	-	.44	.59

and the plot of the probabilities versus the complement anomalies is in Figure 112.

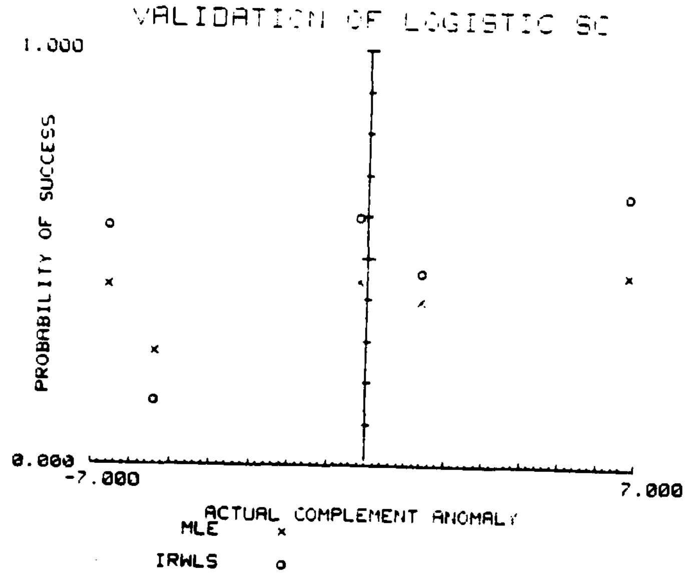


Figure 112. Plot of θ_t versus complement anomalies for maximum likelihood and IRWLS parameters for data set SC

C. DISCUSSION

The validation of the maximum likelihood models for data sets RN and FL appears to be acceptable. However, data set SC does not appear to be acceptably modeled. In fact, as Figure 112 shows, the complement appears to be almost independent of the control. This is also shown by Figure 105 where it can be seen that the contours are very flat and circular about the origin of the (α,β) coordinate system.

The one apparent outlier of data set FL may be viewed as very close, therefore that model can also be assumed to be validated.

VIII. FURTHER FINDINGS

A. SUMMER MONTHS

The further investigation of the summer months, to parallel the modeling of the winter months, yielded some interesting results. These results are shown here with no attempt at analysis.

The summer months appear to be increasing in total rainfall and in variance. This is more true for the Peninsula data sets than for the Valley data set. Figures 113 (RN), 114 (FL), and 115 (SC) show the by-month series of summer months. The total summer rainfall series by year are shown in Figures 116 (RN), 117 (FL), and 118 (SC). The reserved data are not included, yet it can be shown to continue the indicated trends.

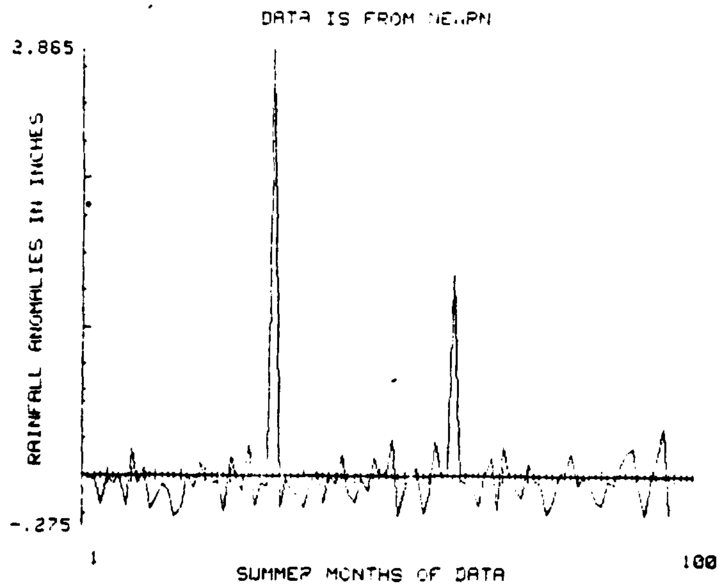


Figure 113. Monthly plot of summer months only, means removed, for data set RN

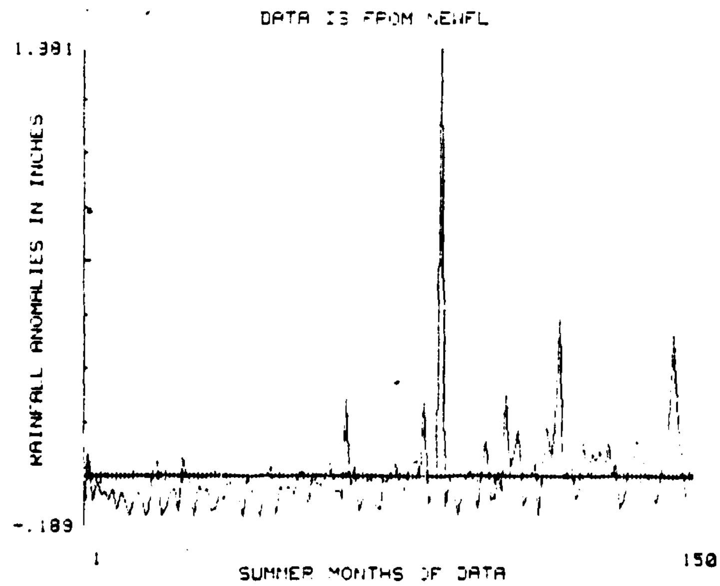


Figure 114. Monthly plot of summer months only, means removed, for data set FL

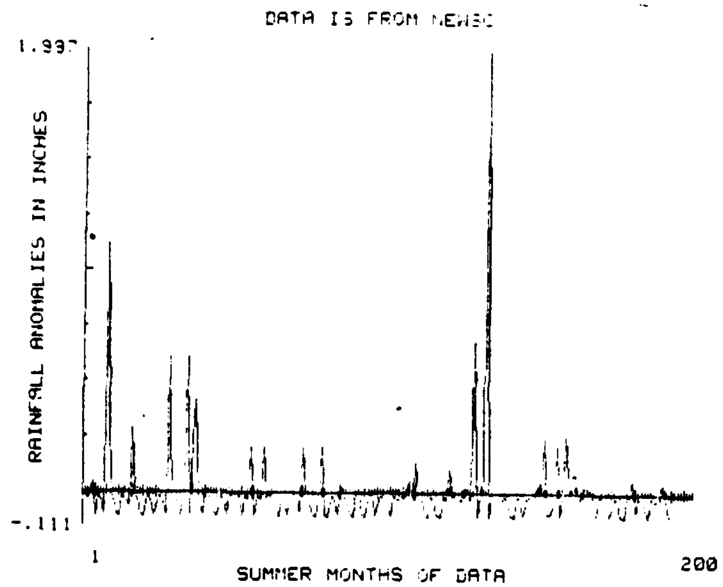


Figure 115. Monthly plot of summer months only, means removed, for data set SC

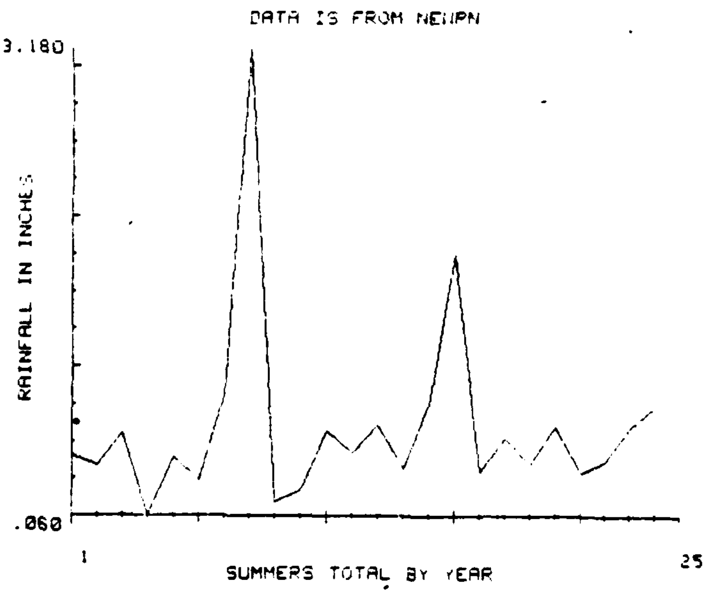


Figure 116. Yearly plot of total summer rainfall for data set RN

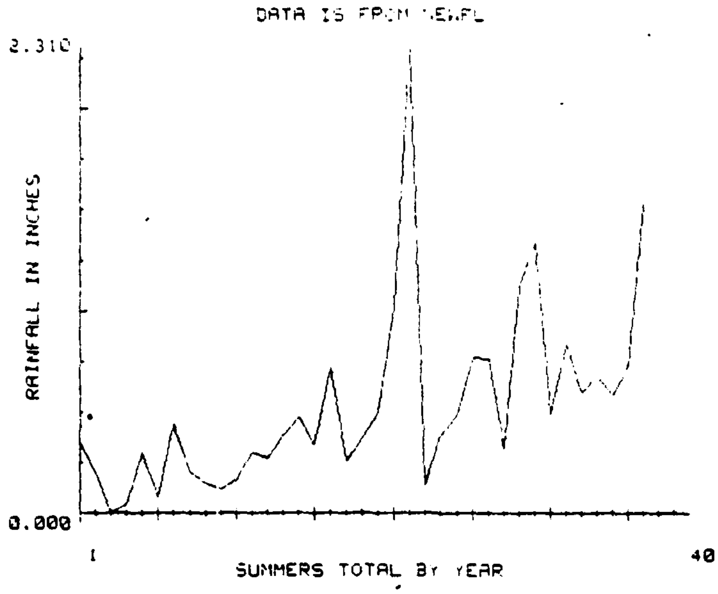


Figure 117. Yearly plot of summer month rainfall for data set FL

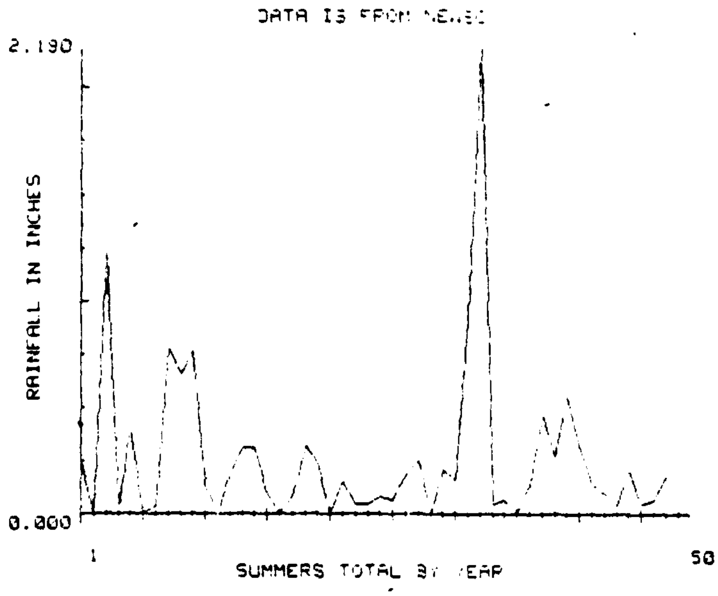


Figure 118. Yearly plot of total summer rainfall for data set SC

A. SIGNIFICANCE OF JANUARY

The identification of January as a possible predictor for its eleven month complement raises further questions. One of the questions is in determining which part of the complement lends the most towards its predictability. Figure 119 is a plot of the log-odds and chi-square statistics for the cumulative complements. Each progressive column, to the right, of Figure 119 indicates these statistics for another cumulated month, i.e., the first column compares the anomalies of January and February by itself, the second column is a comparison of January to February plus March, and so on until the last column is a comparison of January to the entire eleven month complement.

Several occurrences to be noted from the figure are:

(1) The log-odds are consistently greater than zero.

(2) The lack of increased odds and significance during the summer months.

(3) The similarity of RN to FL and their combined difference to SC in the fall.

These indications suggested a further look at January versus the fall months only. This analysis is displayed in Figure 120. The vertical scales of Figure 119 and Figure 120 are the same, yet the horizontal scales differ. This figure has five major divisions. The left-most division looks at January versus singular months in the fall. The second division looks at January versus pairs

of months in the fall, and so on until the right-most column, which is January versus the total fall rainfall. This figure yields no apparent significance, and unstable odds.

The combined information of Figures 119 and 120 are mildly confusing. One possible explanation may be that the summer months somehow cumulate significance and deviation direction, in order to allow the fall contribution. This possible synergistic affect should be explored further.

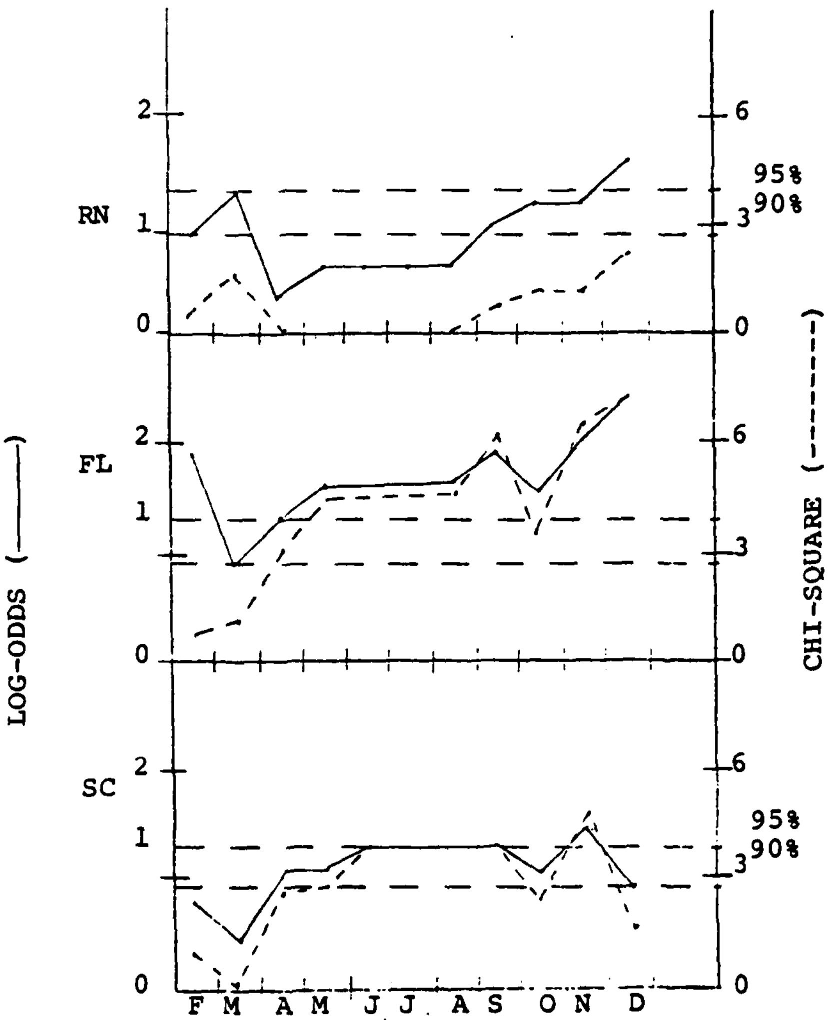


Figure 119. Log-odds and significance versus additional months cumulated through the year

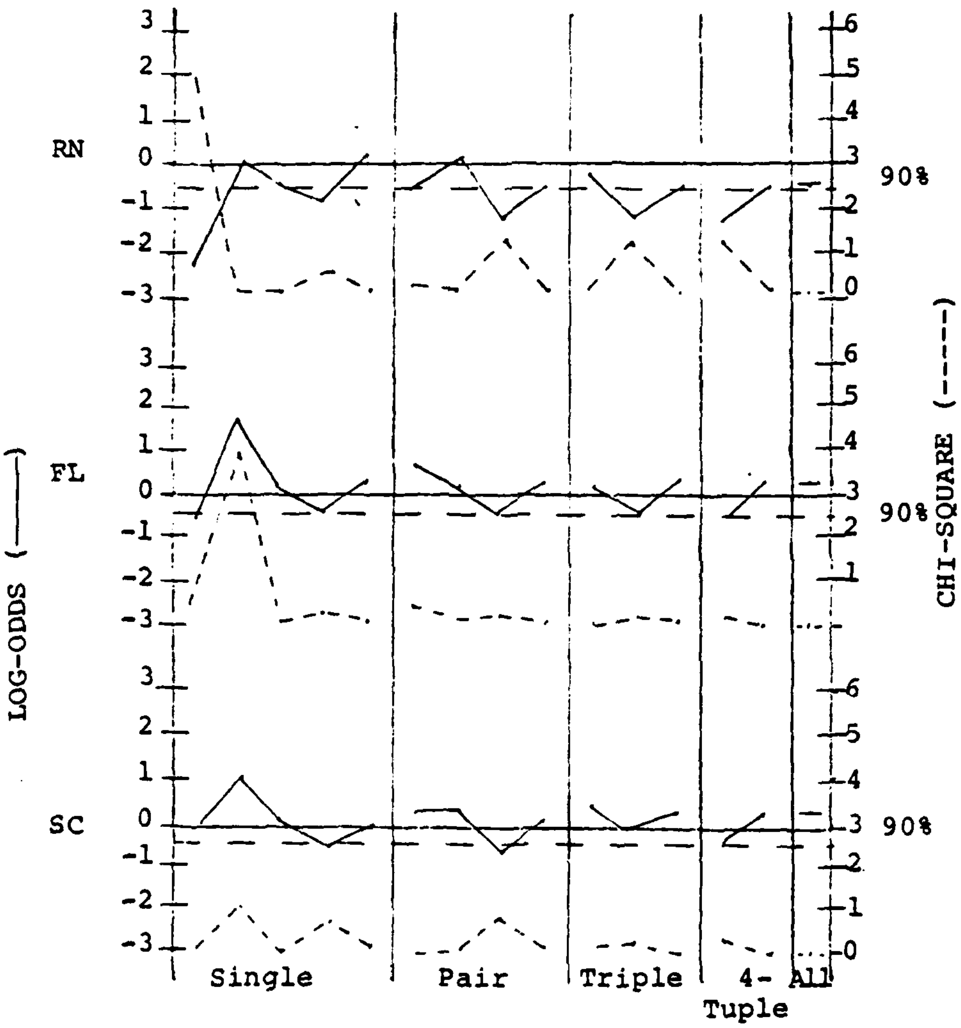


Figure 120. Log-odds and significance versus additional months of the fall

IX. SUMMARY

The analysis of rainfall data is carried out in a comprehensive way. The autoregressive Markovian model of the early sections could not stand up to validation, but it did point to some sort of dichotomy between the seasons.

2x2 contingency analysis was effective in that it brought attention to the predictive ability of January. This identification of January, when followed by the logistic analysis was seen to be successful in two of the three data sets. Thus, the primary conclusion of this thesis is the predictive ability of January rainfall.

The physical reasoning behind this finding must be left to the meteorologist. Further study of the approach used here may lead to improvement in seasonal or annual rainfall forecasts for certain climatic regions.

APPENDIX A

DATA SET RN

RAW

YEAR	OCT	NOV	DEC	JAN	FEB	MAR
	APR	MAY	JUN	JUL	AUG	SEP
1951	.900	2.780	6.880	10.040	2.960	4.410
	1.100	.160	.240	.060	.080	.080
1952	.170	2.190	5.900	2.060	.040	1.200
	1.670	.490	.220	0.000	.110	.070
1953	.380	2.150	.560	4.260	2.260	4.910
	.850	.420	.400	0.000	.160	.050
1954	.030	2.500	3.130	5.820	1.740	.160
	1.740	.870	.060	0.000	0.000	0.000
1955	.080	1.960	9.790	6.090	.060	.150
	1.670	.590	0.000	.070	.020	.360
1956	1.000	0.000	.840	4.650	3.520	1.920
	1.490	2.390	.200	0.000	.070	.030
1957	1.580	.930	3.700	3.710	5.660	7.170
	4.710	.560	.350	.040	0.000	.480
1958	.040	.510	.490	4.850	5.760	.320
	.290	.120	0.000	0.000	.040	3.140
1959	0.000	0.000	.590	4.300	4.530	.840
	.880	.340	0.000	.030	0.000	.130
1960	.070	2.060	.850	1.890	1.170	2.580
	1.290	.720	0.000	0.000	.140	.090
1961	.040	1.740	1.190	2.690	5.170	2.570
	.300	.150	.230	0.000	.250	.150
1962	1.330	.370	2.210	3.050	2.700	4.140
	3.930	.660	.040	.040	.001	.400
1963	1.460	3.770	.530	3.500	.420	2.230
	.220	.860	.220	.090	.350	.010
1964	.780	3.290	6.450	2.560	1.050	2.440
	2.260	.170	.150	.050	.160	.020
1965	.230	6.490	5.560	2.320	1.880	.430
	.270	.130	.120	.280	.090	.320
1966	.090	4.740	4.180	5.290	.450	5.480
	7.110	.400	1.560	.020	.060	.170

YEAR	OCT APR	NOV MAY	DEC JUN	JAN JUL	FEB AUG	MAR SEP
1967	.380	1.610	2.270	3.100	1.400	3.060
	.790	.320	.010	.060	.230	.050
1968	.310	3.130	3.270	9.450	7.310	1.310
	2.700	.120	.420	.040	.001	.120
1969	.500	.720	3.080	5.910	2.040	2.970
	.350	.050	.300	.030	.060	.020
1970	.590	6.170	4.990	1.080	.620	1.960
	1.190	.710	.030	.070	.130	.430
1971	.090	1.990	4.760	1.230	1.050	.030
	.880	.090	.150	.060	.040	.100
1972	2.460	5.950	2.080	6.050	5.880	4.520
	.130	.060	.020	.020	.050	.340
1973	2.200	3.870	4.730	3.730	.910	4.480
	3.400	.030	.370	.250	.020	.010
1974	1.540	.560	2.480	1.340	3.620	4.060
	1.760	.010	.170	.170	.430	.020

Logged Anomalies

YEAR	OCT	NOV	DEC	JAN	FEB	MAR
	APR	MAY	JUN	JUL	AUG	SEP
1951	.204	.237	.753	.863	.272	.556
	-.112	-.170	.040	.005	-.017	-.108
1952	-.281	.063	.620	-.421	-1.065	-.344
	.128	.080	.023	-.054	.010	-.118
1953	-.116	.055	-.367	.121	.078	.644
	-.239	.032	.161	-.054	.054	-.137
1954	-.409	.160	.106	.381	-.096	-.984
	.154	.307	-.117	-.054	-.094	-.185
1955	-.361	-.007	1.067	.420	-1.046	-.993
	.128	.145	-.176	.014	-.074	.122
1956	.255	-1.092	-.702	.193	.404	-.061
	.058	.902	.007	-.054	-.026	-.156
1957	.510	-.435	.236	.011	.792	.968
	.888	.126	.125	-.014	-.094	.207
1958	-.399	-.680	-.913	.228	.807	-.855
	-.599	-.205	-.176	-.054	-.055	1.235
1959	-.438	-1.092	-.348	.129	.606	-.523
	-.223	-.026	-.176	-.024	-.094	-.063
1960	-.370	.026	-.697	-.478	-.329	.143
	-.025	.224	-.176	-.054	.037	-.099
1961	-.399	-.084	-.528	-.233	.715	.140
	-.592	-.179	.031	-.054	.129	-.046
1962	.408	-.778	-.146	-.140	.204	.504
	.741	.188	-.136	-.014	-.093	.151
1963	.462	.470	-.887	-.035	-.754	.040
	-.655	.302	.023	.032	.206	-.175
1964	.139	.364	.696	-.269	-.386	.103
	.328	-.162	-.036	-.005	.054	-.166
1965	-.231	.921	.569	-.339	-.046	-.775
	-.615	-.196	-.062	.193	-.008	.092
1966	-.352	.655	.333	.300	-.733	.736
	1.239	.018	.764	-.034	-.036	-.028

YEAR	OCT	NOV	DEC	JAN	FEB	MAR
	APR	MAY	JUN	JUL	AUG	SEP
1967	-.116	-.133	-.127	-.128	-.229	.269
	-.272	-.041	-.166	.005	.113	-.137
1968	-.168	.326	.140	.808	1.013	-.295
	.454	-.205	.175	-.014	-.093	-.072
1969	-.033	-.550	.094	.394	.008	.246
	-.554	-.270	.087	-.024	-.036	-.166
1970	.026	.878	.478	-.807	-.622	-.047
	-.070	.218	-.146	.014	.028	.172
1971	-.352	.003	.439	-.737	-.386	-1.103
	-.223	-.232	-.036	.005	-.055	-.090
1972	.803	.846	-.187	.414	.824	.576
	-.732	-.260	-.156	-.034	-.045	.107
1973	.725	.491	.434	.015	-.457	.568
	.628	-.289	.139	.169	-.074	-.175
1974	.494	-.648	-.065	-.689	.426	.489
	.161	-.309	-.019	.103	.264	-.166

Reserved Data

YEAR	OCT APR	NOV MAY	DEC JUN	JAN JUL	FEB AUG	MAR SEP
1974	1.540 1.760	.560 .010	2.480 .170	1.340 .170	3.620 .430	4.060 .020
1975	1.700 1.740	.520 .070	.370 .170	.180 .020	2.970 .970	1.520 .420
1976	.600 .040	.720 1.210	2.080 .080	1.740 .030	.830 .020	1.750 .650
1977	.140 5.430	.540 .020	5.850 .080	6.780 .040	4.780 .001	5.240 .290
1978	.020 .580	2.130 .290	1.590 .020	4.820 .350	4.520 .090	4.410 .020
1979	1.800 1.770	2.850 .570	3.180 .040	5.950 .730	4.780 .090	2.400 .090

APPENDIX B

DATA SET FL

RAW

YEAR	OCT APR	NOV MAY	DEC JUN	JAN JUL	FEB AUG	MAR SEP
1937	.640 2.250	1.270 .050	4.600 0.000	4.990 .180	7.590 0.000	5.630 .170
1938	1.280 .280	.770 .630	2.090 .070	3.190 0.000	1.620 0.000	2.370 .130
1939	.750 .720	.480 .130	1.470 0.000	8.430 0.000	8.200 0.000	4.240 0.000
1940	0.000 5.760	.790 .470	4.520 .040	6.150 0.000	7.730 0.000	5.500 0.000
1941	.800 3.660	.280 .790	7.870 0.000	3.480 .120	2.910 .180	2.430 0.000
1942	1.130 1.300	1.470 .100	.970 0.000	3.050 0.000	2.750 .060	4.240 .020
1943	.620 1.160	.420 .830	3.830 .240	3.690 .100	8.080 .100	.710 0.000
1944	1.130 .430	4.650 .300	1.940 .110	.980 0.000	3.180 .060	2.830 .030
1945	1.370 .040	1.400 .640	4.230 0.000	1.080 0.000	2.560 0.000	3.000 .150
1946	.260 .330	4.310 .210	1.970 .120	.440 0.000	1.130 0.000	1.950 0.000
1947	1.160 3.250	.470 .400	1.940 .170	.090 0.000	2.140 0.000	4.610 0.000
1948	2.130 .080	.350 .420	3.200 .050	1.700 .080	2.840 .170	4.730 0.000
1949	.030 1.470	1.690 .190	1.370 .110	3.120 0.000	1.740 .070	1.640 .090
1950	1.760 1.060	3.000 .270	2.370 .120	2.060 .110	2.190 .100	1.480 .050
1951	.670 1.100	3.350 .120	6.040 .170	8.010 .040	2.110 .070	4.550 .190
1952	.140 1.250	2.210 .570	5.040 .220	1.710 0.000	0.000 .050	.770 .070

AD-A110 816 NAVAL POSTGRADUATE SCHOOL MONTEREY CA F/G 4/2

A STATISTICAL ANALYSIS OF MONTHLY RAINFALL FOR MONTEREY PENINSU--ETC(U)

MAR 81 D F DAVIS

UNCLASSIFIED NL

3 of 3

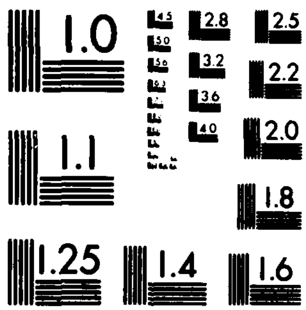
AD A110816

END

DATE FILMED

3 82

DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963-A

YEAR	OCT APR	NOV MAY	DEC JUN	JAN JUL	FEB AUG	MAR SEP
1953	.280	2.050	.310	3.330	1.580	4.410
	.780	.390	.510	.020	.160	.030
1954	.020	2.170	2.820	4.560	1.560	.130
	1.660	.960	.050	.060	.100	.050
1955	.040	1.650	8.310	4.820	1.740	.180
	1.560	.590	0.000	.060	.090	.220
1956	.650	0.000	.770	4.320	2.610	1.890
	1.340	2.640	.210	.020	.050	.220
1957	2.020	.850	3.380	2.620	4.620	6.760
	3.980	.490	.220	.150	.100	.550
1958	.040	.390	.400	4.470	5.630	.340
	.460	.120	.020	.030	.090	2.170
1959	.030	0.000	.620	3.990	3.270	.650
	.870	.410	.010	.080	.040	.010
1960	.070	1.450	.640	1.840	1.110	2.420
	1.200	.750	.040	.040	.160	.140
1961	.100	2.030	1.380	2.230	5.830	2.400
	.270	.150	0.000	.130	.280	.080
1962	1.360	.440	2.260	3.830	2.270	4.320
	3.320	.650	.050	.110	.040	.570
1963	1.360	5.260	.540	3.190	.390	2.300
	.100	.690	.190	.190	.330	.050
1964	1.060	3.160	5.970	2.370	.930	2.430
	2.180	.140	.060	.060	.190	.010
1965	.160	6.250	6.230	3.050	1.490	.510
	.170	.140	.190	.320	.200	.410
1966	.140	4.690	3.880	5.750	.490	5.070
	6.810	.790	.890	.060	.160	.230
1967	.270	1.810	2.060	3.110	1.350	3.180
	.720	.290	.040	0.000	.270	.180
1968	.430	3.300	2.750	8.480	7.960	1.200
	2.490	.190	.260	.150	.240	.190

YEAR	OCT APR	NOV MAY	DEC JUN	JAN JUL	FEB AUG	MAR SEP
1969	.410 .430	1.010 .080	2.800 .320	5.550 .050	3.020 .190	1.860 .040
1970	.540 1.190	7.440 .580	4.450 .050	.370 .110	.640 .170	1.940 .350
1971	.170 .720	1.550 .180	4.680 .290	1.000 .090	.440 .100	.080 .100
1972	2.540 .310	5.520 .160	1.850 .030	5.550 .070	6.070 .220	3.530 .410
1973	1.960 2.990	4.750 .110	3.270 .810	3.760 .410	1.190 .210	4.500 .100

Logged Anomalies

YEAR	OCT APR	NOV MAY	DEC JUN	JAN JUL	FEB AUG	MAR SEP
1937	.011	-.187	.454	.392	.932	.708
	.383	-.289	-.130	.095	-.106	.011
1938	.340	-.436	-.141	.035	-.255	.031
	-.549	.151	-.062	-.071	-.106	-.024
1939	.076	-.615	-.365	.846	1.001	.473
	-.254	-.216	-.130	-.071	-.106	-.146
1940	-.484	-.425	.439	.569	.948	.688
	1.115	.047	-.091	-.071	-.106	-.146
1941	.104	-.760	.914	.102	.145	.049
	.743	.244	-.130	.042	.060	-.146
1942	.272	-.103	-.591	.001	.103	.473
	.037	-.243	-.130	-.071	-.047	-.126
1943	-.001	-.656	.306	.148	.988	-.647
	-.026	.266	.085	.024	-.010	-.146
1944	.272	.725	-.191	-.715	.212	.159
	-.438	-.076	-.025	-.071	-.047	-.117
1945	.379	-.132	.385	-.666	.051	.203
	-.757	.157	-.130	-.071	-.106	-.006
1946	-.253	.663	-.181	-1.033	-.462	-.102
	-.511	-.147	-.017	-.071	-.106	-.146
1947	.286	-.622	-.191	-1.312	-.074	.541
	.651	-.001	.027	-.071	-.106	-.146
1948	.657	-.707	.166	-.405	.127	.562
	-.719	.013	-.001	.006	.051	-.146
1949	-.454	-.018	-.406	.018	-.210	-.213
	.100	-.164	-.025	-.071	-.038	-.060
1950	.532	.379	-.054	-.279	-.058	-.275
	-.073	-.099	-.017	.034	-.010	-.097
1951	.029	.463	.682	.000	-.084	.530
	-.054	-.225	.027	-.032	-.038	.020
1952	-.353	.159	.529	-.401	-1.218	-.613
	.015	.113	.069	-.071	-.057	-.070

YEAR	OCT APR	NOV MAY	DEC JUN	JAN JUL	FEB AUG	MAR SEP
1953	-.237	.108	-.999	.068	-.271	.505
	-.219	-.009	.282	-.051	.043	-.117
1954	-.464	.147	.071	.318	-.278	-1.061
	.182	.335	-.081	-.013	-.010	-.097
1955	-.444	-.032	.962	.363	-.210	-1.018
	.144	.126	-.130	-.013	-.019	.053
1956	.017	-1.007	-.698	.274	.065	-.122
	.054	.954	.061	-.051	-.057	.053
1957	.622	-.392	.208	-.111	.508	.865
	.810	.061	.069	.069	-.010	.292
1958	-.444	-.678	-.933	.301	.673	-.891
	-.417	-.225	-.110	-.041	-.019	1.009
1959	-.454	-1.007	-.787	.210	.233	-.683
	-.170	.006	-.120	.006	-.066	-.136
1960	-.416	-.111	-.774	-.354	-.472	.046
	-.007	.222	-.091	-.032	.043	-.015
1961	-.388	.102	-.402	-.225	.703	.040
	-.557	-.198	-.130	.051	.141	-.069
1962	.375	-.642	-.087	.177	-.034	.488
	.667	.163	-.081	.034	-.066	.305
1963	.375	.827	-.837	.035	-.889	.010
	-.701	.187	.044	.103	.180	-.097
1964	.239	.418	.673	-.183	-.561	.049
	.361	-.207	-.072	-.013	.068	-.136
1965	-.335	.974	.709	.001	-.306	-.771
	-.639	-.207	.044	.207	.077	.198
1966	-.353	.732	.316	.512	-.820	.620
	1.260	.244	.507	-.013	.043	.061
1967	-.245	.026	-.151	.016	-.364	.247
	-.254	-.083	-.091	-.071	.133	.019
1968	-.126	.452	.053	.051	.974	-.395
	.454	-.164	.101	.069	.110	.028

YEAR	OCT APR	NOV MAY	DEC JUN	JAN JUL	FEB AUG	MAR SEP
1969	-.140 -.438	-.309 -.261	.066 .148	.482 -.022	.173 .068	-.133 -.107
1970	-.052 -.012	1.126 .120	.427 -.081	-.772 .034	-.724 .051	-.105 .154
1971	-.327 -.254	-.071 -.172	.468 .125	-.705 .015	-.854 -.010	-1.107 -.051
1972	.780 -.526	.868 -.189	-.222 -.100	.482 -.003	.738 .093	.327 .198
1973	.602 .588	.742 -.234	.183 .463	.162 .273	-.434 .085	.521 -.051

Reserved Data

YEAR	OCT APR	NOV MAY	DEC JUN	JAN JUL	FEB AUG	MAR SEP
1974	1.960 1.580	.720 .110	2.430 .180	1.700 .380	2.760 .450	4.680 .080
1975	1.480 1.660	.470 .070	.410 .130	.200 .140	2.820 1.330	1.060 .820
1976	.110 .030	.780 .960	1.660 .110	1.700 .020	.630 .040	1.900 1.050
1977	.170 5.210	.620 .030	5.570 .100	5.780 .090	4.710 .050	5.160 .310
1978	.100 .560	2.220 .130	1.340 .010	4.700 .330	3.650 .120	3.710 .040
1979	1.580 1.520	3.360 .580	3.830 .040	5.120 .900	4.420 .140	2.030 .130

APPENDIX C

DATA SET SC

RAW

YEAR	OCT	NOV	DEC	JAN	FEB	MAR
	APR	MAY	JUN	JUL	AUG	SEP
1926	.420	8.360	1.640	2.440	7.580	1.390
	1.070	.280	.100	0.000	0.000	.150
1927	.800	2.100	2.830	.880	2.430	3.580
	.960	.050	0.000	0.000	0.000	0.000
1928	.020	3.120	3.410	1.260	1.370	2.890
	1.290	0.000	1.230	0.000	0.000	0.000
1929	0.000	0.000	.830	5.370	3.260	4.250
	1.330	1.610	0.000	0.000	0.000	.030
1930	.050	1.480	.050	4.260	1.560	.920
	.490	.660	.400	0.000	0.000	0.000
1931	.020	1.750	10.260	4.320	4.690	.800
	.140	.360	0.000	0.000	0.000	0.000
1932	0.000	.180	3.120	6.920	.920	1.570
	.330	.820	.040	0.000	0.000	0.000
1933	1.120	0.000	8.250	3.150	5.290	0.000
	.130	.850	.720	0.000	0.000	.060
1934	.150	2.890	2.690	7.000	.770	4.880
	5.910	0.000	0.000	0.000	.650	0.000
1935	.370	.810	2.010	2.660	9.970	1.680
	2.460	.480	.340	.430	0.000	0.000
1936	.430	0.000	4.160	4.620	6.770	7.790
	.520	0.000	.130	0.000	0.000	0.000
1937	.060	1.010	6.590	3.440	13.020	3.090
	2.630	0.000	0.000	0.000	0.000	0.000
1938	.810	1.040	2.480	3.540	2.700	3.080
	.420	.220	.100	0.000	0.000	.100
1939	1.040	.370	2.380	9.200	8.960	2.320
	.600	.270	0.000	0.000	0.000	.310
1940	1.040	.370	2.380	9.200	8.960	2.320
	.600	.270	0.000	0.000	0.000	.310
1941	.450	.330	9.450	5.180	10.420	8.920
	5.000	.610	.100	0.000	0.000	0.000

YEAR	OCT APR	NOV MAY	DEC JUN	JAN JUL	FEB AUG	MAR SEP
1942	1.130	1.040	10.490	5.710	2.460	3.280
	4.910	1.010	0.000	0.000	0.000	0.000
1943	.740	1.960	1.630	8.400	2.420	3.740
	1.070	0.000	.090	0.000	0.000	0.000
1944	.740	.360	3.330	3.940	9.280	1.000
	1.940	.750	.320	0.000	0.000	0.000
1945	1.080	3.240	2.390	1.330	7.610	4.880
	.280	.200	0.000	0.000	.240	0.000
1946	2.800	2.050	7.690	.850	2.770	3.280
	0.000	.430	0.000	0.000	0.000	0.000
1947	.790	.670	2.360	.050	2.480	4.380
	3.860	.620	.150	0.000	0.000	6.000
1948	1.740	.150	6.010	1.180	3.040	5.960
	.150	.510	0.000	.020	.030	0.000
1949	.190	1.470	1.690	6.090	2.970	2.230
	1.470	.260	0.000	0.000	.050	0.000
1950	2.720	6.330	3.040	2.950	1.940	1.840
	1.370	.690	.050	0.000	0.000	.030
1951	1.440	4.010	8.360	9.800	1.610	6.820
	.800	.220	0.000	0.000	0.000	.070
1952	.070	3.000	8.760	2.680	0.000	2.090
	1.890	.620	.100	0.000	.090	0.000
1953	.300	2.160	.490	4.370	3.280	4.680
	.530	.190	.250	0.000	0.000	0.000
1954	0.000	1.980	2.980	5.530	1.990	.310
	2.840	1.360	0.000	0.000	0.000	0.000
1955	.030	2.240	14.940	6.590	2.250	.510
	1.600	.480	0.000	0.000	0.000	.210
1956	.790	.020	.530	5.210	5.110	1.570
	1.720	1.930	.090	0.000	0.000	.070
1957	1.470	.930	3.990	4.630	9.680	7.820
	0.500	.560	.140	0.000	0.000	.780

YEAR	OCT APR	NOV MAY	DEC JUN	JAN JUL	FEB AUG	MAR SEP
1958	.030	1.010	.270	6.350	6.920	.230
	.110	.090	0.000	0.000	.090	2.100
1959	0.000	0.000	.710	5.260	6.520	.740
	1.910	.330	0.000	0.000	0.000	.050
1960	.050	3.860	1.810	1.910	.910	2.490
	.860	.520	.070	0.000	0.000	0.000
1961	.170	2.620	1.300	2.000	11.290	2.960
	.110	.340	0.000	0.000	0.000	0.000
1962	1.810	.100	2.290	7.810	2.200	4.420
	3.990	.230	.050	0.000	0.000	.090
1963	1.300	3.640	.460	4.000	.460	3.080
	.390	1.350	.170	0.000	.300	0.000
1964	1.290	3.500	4.830	2.830	1.050	2.520
	2.850	.140	.020	0.000	.250	0.000
1965	.070	6.630	4.790	1.810	1.340	.680
	.270	0.000	.080	.290	0.000	.180
1966	0.000	3.740	5.410	6.090	.430	6.160
	6.410	.520	.190	0.000	0.000	.130
1967	.270	1.750	1.870	3.360	1.010	3.290
	.670	.430	.120	0.000	.020	0.000
1968	.330	1.950	4.060	15.160	11.970	1.020
	2.140	.120	.100	0.000	0.000	0.000
1969	.340	.790	3.110	7.000	4.610	1.510
	.810	.070	.040	0.000	0.000	0.000
1970	.110	5.860	6.440	1.200	.530	1.490
	1.210	.150	0.000	0.000	.040	.170
1971	.240	1.430	5.560	1.190	.980	.010
	.090	.080	.020	0.000	0.000	.030
1972	2.850	5.790	2.660	7.590	9.200	4.430
	.100	.020	0.000	0.000	0.000	.070
1973	1.810	6.310	2.360	4.440	1.980	4.740
	1.970	0.000	.110	.070	0.000	0.000

Logged Anomalies

YEAR	OCT / APR	NOV / MAY	DEC / JUN	JAN / JUL	FEB / AUG	MAR / SEP
1926	-.093	1.275	-.437	-.347	.705	-.372
	-.090	-.073	.003	-.015	-.031	.065
1927	.144	.170	-.065	-.952	-.212	.279
	-.144	-.271	-.092	-.015	-.031	-.075
1928	-.424	.455	.076	-.768	-.581	.115
	.011	-.320	.710	-.015	-.031	-.075
1929	-.444	-.961	-.804	.269	.005	.415
	.029	.639	-.092	-.015	-.031	-.046
1930	-.395	-.053	-1.359	.077	-.504	-.591
	-.418	.187	.244	-.015	-.031	-.075
1931	-.424	.050	1.013	.089	.294	-.655
	-.686	-.013	-.092	-.015	-.031	-.075
1932	-.444	-.796	.008	.486	-.792	-.299
	-.532	.279	-.053	-.015	-.031	-.075
1933	.308	-.961	.817	-.160	.395	-1.243
	-.695	.295	.450	-.015	-.031	-.017
1934	-.304	.397	-.102	.496	-.873	.529
	1.116	-.320	-.092	-.015	.469	-.075
1935	-.129	-.368	-.306	-.285	.951	-.257
	.424	.072	.200	.343	-.031	-.075
1936	-.086	-.961	.233	.143	.606	.929
	-.398	-.320	.030	-.015	-.031	-.075
1937	-.386	-.263	.619	-.092	1.196	.964
	.472	-.320	-.092	-.015	-.031	-.075
1938	.150	-.248	-.161	-.070	-.136	.163
	-.466	-.121	.003	-.015	-.031	.020
1939	.269	-.646	-.190	.739	.854	-.043
	-.347	-.081	-.092	-.015	-.031	.195
1940	.269	-.646	-.190	.739	.854	-.043
	-.347	-.081	-.092	-.015	-.031	.195
1941	-.072	-.676	.930	.230	.991	1.052
	.908	.156	.003	-.015	-.031	-.075

YEAR	OCT APR	NOV MAY	DEC JUN	JAN JUL	FEB AUG	MAR SEP
1942	.312	-.248	1.033	.321	-.203	.211
	.960	.378	-.092	-.015	-.031	-.075
1943	.110	.124	-.441	.658	-.215	.313
	-.090	-.320	-.006	-.015	-.031	-.075
1944	.110	-.654	.057	.014	.886	-.550
	.261	.239	.185	-.015	-.031	-.075
1945	.289	.483	-.187	-.737	.709	.529
	-.570	-.138	-.092	-.015	.184	-.075
1946	.891	.154	.754	-.968	-.117	.211
	-.017	.037	-.092	-.015	-.031	-.075
1947	.138	-.448	-.196	-1.534	-.197	.440
	.764	.162	.048	-.015	-.031	-.075
1948	.364	-.821	.539	-.804	-.048	.697
	-.677	.092	-.092	.005	-.002	-.075
1949	-.270	-.057	-.419	.376	-.066	-.071
	.087	-.089	-.092	-.015	.017	-.075
1950	.870	1.031	-.012	-.209	-.366	-.530
	.046	.204	-.043	-.015	-.031	-.046
1951	.448	.650	.828	.797	-.485	.814
	-.229	-.121	-.092	-.015	-.031	-.007
1952	-.376	.425	.870	-.280	-1.444	-.115
	.244	.162	.003	-.015	.055	-.075
1953	-.181	.189	-1.009	.098	.010	.494
	-.392	-.146	.131	-.015	-.031	-.075
1954	-.444	.131	-.027	.293	-.349	-.973
	.528	.538	-.092	-.015	-.031	-.075
1955	-.414	.214	1.361	.444	-.266	-.831
	.130	.072	-.092	-.015	-.031	.116
1956	.130	-.941	-.983	.243	.366	-.299
	.183	.755	-.006	-.015	-.031	-.007
1957	.460	-.304	.199	.145	.924	.934
	1.474	.124	.039	-.015	-.031	.502

YEAR	OCT APR	NOV MAY	DEC JUN	JAN JUL	FEB AUG	MAR SEP
1958	-.414	-.263	-1.169	.412	.625	-1.036
	-.713	-.234	-.092	-.015	.055	1.056
1959	-.444	-.961	-.872	.251	.573	-.689
	.251	-.035	-.092	-.015	-.031	-.026
1960	-.395	.620	-.375	-.515	-.797	.007
	-.197	.098	-.025	-.015	-.031	-.075
1961	-.287	.325	-.575	-.484	1.064	.133
	-.713	-.028	-.092	-.015	-.031	-.075
1962	.589	-.866	-.217	.593	-.281	.447
	.790	-.113	-.043	-.015	-.031	.011
1963	.389	.574	-1.030	.026	-1.066	.163
	-.488	.534	.065	-.015	.231	-.075
1964	.385	.543	.355	-.240	-.726	.015
	.531	-.189	-.072	-.015	.192	-.075
1965	-.376	1.071	.348	-.550	-.594	-.724
	-.578	-.320	-.015	.240	-.031	.090
1966	-.444	.595	.450	.376	-1.087	.726
	1.186	.098	.082	-.015	-.031	.047
1967	-.205	.050	-.354	-.110	-.746	.213
	-.304	.037	.021	-.015	-.012	-.075
1968	-.159	.121	.213	1.200	1.118	-.540
	.327	-.207	.003	-.015	-.031	-.075
1969	-.151	-.379	.005	.496	.280	-.323
	-.224	-.253	-.053	-.015	-.031	-.075
1970	-.339	.965	.599	-.794	-1.019	-.331
	-.024	-.180	-.092	-.015	.008	.082
1971	-.229	-.073	.473	-.799	-.761	-1.233
	-.181	-.243	-.072	-.015	-.031	-.046
1972	.904	.954	-.111	.568	.878	.449
	-.722	-.300	-.092	-.015	-.031	-.007
1973	.589	1.028	-.196	.111	-.352	.504
	.271	-.320	.012	.053	-.031	-.075

Reserved Data

YEAR	OCT APR	NOV MAY	DEC JUN	JAN JUL	FEB AUG	MAR SEP
1974	1.840 1.660	.760 .020	3.570 .040	2.040 .070	5.490 .120	8.200 0.000
1975	1.310 1.460	.730 .050	.330 .180	.080 .001	2.350 1.050	1.450 .750
1976	1.450 .140	.600 .970	1.850 .080	2.450 0.000	.460 0.000	1.800 .480
1977	.020 3.600	.400 .030	5.680 0.000	9.990 0.000	8.400 0.000	7.070 .360
1978	0.000 .380	2.720 .260	1.200 0.000	5.810 .120	4.660 0.000	3.920 0.000
1979	1.010 2.550	2.140 .420	3.870 0.000	5.220 .950	10.890 0.000	2.890 0.000

LIST OF REFERENCES

1. Box, E.P., and Jenkins, G.M., Time Series Analysis: Forecasting and Control, 2d. ed., Holden-Day, 1970.

2. Gaver, D.P., Logistic Analysis, class presentation notes for OA4910, Naval Postgraduate School, Fall 1980.

3. Fleiss, J.L., Statistical Methods for Rates and Proportions, Wiley, 1973.

4. Fisher, R.A., The Design of Experiments, 7th. ed., Oliver and Boyd, 1960.

5. Dixon, W.J., and Massey, F.J. Jr., Introduction to Statistical Analysis, 3d ed., McGraw-Hill, 1969.

6. Brownlee, K.A., Statistical Theory and Methodology in Science and Engineering, 2d. ed., Wiley, 1965.

7. Mood, A.M., Grayhill, F.A., and Boes, D.C., Introduction to the Theory of Statistics, 3d ed., McGraw-Hill, 1963.

8. Nelson, C.R., Applied Time Series Analysis for Managerial Forecasting, Holden-Day, 1973.

9. Pierce, D.A., "A Survey of Recent Developments in Seasonal Adjustment", The American Statistician, Vol. 34, No. 3, August 1980.

10. Kilmartin, R.F., and Peterson, J.R., "Rainfall-Runoff Regression with Logarithmic Transforms and Zeros in the Data", Water Resources Research, Vol. 8, No. 4., August 1972.

11. Hipel, K.W., BcBean, E.A., and McLeod, A.I., "Hydrologic Generating Model Selection", Journal of the Water Resources Planning and Mangement Division, W.R. 2, September 1979.

12. Naval Postgraduate School Report NPS55-78-034, An Interactive Software Package for Time Series Analysis, by F. Russell Richards and Stephen R. Woodall, November 1978.

13. Mosteller, F., and Tukey, J.W., Data Analysis and Regression; A Second Course in Statistics, Addison-Wesley, 1977.

14. Kendall, M., Time-Series, 2d. ed., Hafner Press, 1976.

15. McNeil, D.R., Interactive Data Analysis, A Practical Primer, Wiley, 1977.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Technical Information Center Cameron Station Alexandria, Virginia 22314	2
2.	Library, Code 0142 Naval Postgraduate School Monterey, California 93940	2
3.	Professor D.P. Gaver, Code 55Gv Naval Postgraduate School Monterey, California 93940	1
4.	Professor P.A. Jacobs, Code 55Jc Naval Postgraduate School Monterey, California 93940	1
5.	Professor R.J. Renard, Code 63Rd Naval Postgraduate School Monterey, California 93940	2
6.	Monterey Water Management District Attn: Mr. Bruce Buel P.O. Box 85 Monterey, California 93940	2
7.	Professor F.D. Faulkner, Code 53Fa Naval Postgraduate School Monterey, California 93940	1
8.	Professor A.L. Schoenstadt, Code 53Zh Naval Postgraduate School Monterey, California 93940	1
9.	Professor K.T. Marshall, Code 55Mt Naval Postgraduate School Monterey, California 93940	1
10.	Captain David F. Davis, USA 17 Glenmore Rd. Pueblo, Colorado 81001	1

END

DATE
FILMED
3-82

DTIC